GLOBAL ANALYSIS OF SUMO-BINDING PROTEINS IDENTIFIES SUMOYLATION AS A KEY REGULATOR OF THE INO80 CHROMATIN REMODELING COMPLEX

by

Eric Cox

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

September, 2014

Abstract

The functional protein microarray is a powerful and versatile systems biology and proteomics tool that allows the rapid activity profiling of thousands of proteins in parallel. Applications of functional protein microarrays range from the identification of protein-binding properties, to surveying targets of posttranslational modifications, to uncovering novel enzymatic activities. Since the development of the yeast proteome microarray over 10 years ago [1], more recent work has seen the development of complete and near-complete proteome arrays representing viruses, bacteria and plants [2-4]. However, most existing human protein microarrays are comprised of only a minority of the human proteome [5-9]. We have recently developed a human proteome microarray, the HuProt array, which includes nearly 20,000 full-length human proteins [10].

SUMOylation is an essential posttranslational modification in most organisms that is thought to function through its ability to modulate the protein-protein interactions of a SUMO target protein. Accordingly, the function of SUMOylation can be better understood through the identification of SUMO-modified targets as well as downstream SUMO-interacting proteins. Recently, we have conducted SUMOylation assays using the HuProt microarray to identify numerous previously uncharacterized SUMO E3 ligase-dependent substrates using a subset of human SUMO E3 ligases. In order to identify novel SUMO-interacting proteins, we developed a SUMO-binding assay using the human proteome microarray. We then integrated SUMO-binding and SUMOylation data, as well as protein-protein interaction data from publicly available databases to perform network motif analysis. We focused on a single network motif we termed a SUMOmod PPI (SUMO-modulated Protein-Protein Interaction) that included the INO80 chromatin remodeling complex subunits

TFPT and INO80E. We validated the SUMO-binding activity of INO80E and that TFPT is a SUMO substrate both *in vitro* and *in vivo.* We then went on to demonstrate a key role for SUMOylation in mediating the interaction between these two proteins, both *in vitro* and *in vivo*. By demonstrating a key role for SUMOylation in regulating the INO80 chromatin remodeling complex, this work illustrates the power of bioinformatics analysis of large datasets in predicting novel biological phenomena.

Thesis Advisors:

Dr. Seth Blackshaw

Associate Professor, The Solomon H. Snyder Department of Neuroscience

Johns Hopkins University School of Medicine

Dr. Heng Zhu

Associate Professor, Department of Pharmacology and Molecular Sciences

Johns Hopkins University School of Medicine

Thesis Reader:

Dr. Michael Matunis

Professor, Department of Biochemistry and Molecular Biology

Bloomberg School of Public Health, Johns Hopkins University

Dedication

I would like to dedicate this dissertation to my wife, Annette, for always supporting me, inspiring me with her sense of adventure, and making life so much fun.

Acknowledgements

I would like to thank my Dad, my first science teacher who helped me to design science fair projects in the garage or at his shop, measuring the voltage output of a homemade penny-vinegar battery or the heat output of a Duraflame log. I would like to thank my Mom for always being my cheerleader, and being more excited about my graduation than me. I would like to thank my stepmom, Christy, for reminding me not to take myself too seriously.

I would like to thank Seth and Heng, the two most enthusiastic scientists I have ever known or seen. I can still remember during my rotation in Seth's lab, when Seth yelled across the room to announce that I had purified a protein, and I remember feeling like I had won the Nobel Prize. Or when Heng and Seth toasted Ijee and I with champagne when we completed our SUMOylation screen. Or when Seth or Heng are excited about a new project or idea, how the volume in their voices gets louder and louder and sometimes you might think something terrible has happened, but it's okay, they're just so excited about science! I feel very lucky to have had the opportunity to train with two such accomplished scientists that are also great guys to have a beer with, and are willing to make themselves available to do just that.

I would like to thank Jun Seop, for asking me so many questions, which sometime annoyed me at the time, but always forced me to think a little bit harder and more carefully about the experiments that I was doing.

I also want to thank Hee-sool and Jimmy, who I got to know much better towards the end of grad school for their friendship and advice and for always being there to talk to.

Thanks to my lab partner Ijee for being a great friend and coworker, and for filling the entire Hit center with her laughter, a laugh that echoes off the walls on the other side of the hit center two labs away, and always made me smile and seemed to lighten the mood of the entire floor. Grad school isn't that hard when you have friends like Ijee.

Thanks to Mike Matunis, for teaching me and our labs so much about SUMO and the field, giving advice on my (various) projects, encouraging me to apply for the AHA fellowship, and being my thesis reader. Thanks to my other thesis committee members, Jiang Qian and Jeremy Nathans for sticking with me, listening to my ideas, and providing lots of helpful ones of their own.

I also want to thank all of my other Baltimore scientist friends, especially Ian, Tarran, Brad, Sean, Jota, Korin, and Maria, who were all grad students themselves, for all their pep talks and advice and encouragement. It would have been impossible without their support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction to Protein Microarray Technology

This section has been adapted from: Zhu, H., Cox, E., and Qian, J. (2012) The Functional Protein Microarray as Molecular Decathlete: A Versatile Player in Clinical Proteomics. *Proteomics Clin Appli.* 6, 548-562

1.1. Introduction

The concept of microarrays was developed from an earlier concept termed ambient analyte immunoassay, first introduced by Roger Ekins in 1989. In the following decade, microarrays were first successfully realized as DNA or oligonucleotide microarrays, which allowed the quantification of the mRNA expression levels of thousands of genes in parallel. This technology has changed many aspects of biological research. Though extremely successful, the chemistry of DNA hybridization precludes its application for studying proteins, which are considered the major driving force in cells. Consistent with this view, mRNA profiles do not always correlate with protein expression as reported in many recent mass spectrometry studies [11-13]. Therefore, protein microarrays were developed as a high-throughput tool to overcome the limitations of DNA microarrays, and to provide a versatile platform for protein functional analyses [8, 14, 15].

At the beginning of the development of protein array technology, bacterial strains of a cDNA expression library were gridded and grown on nylon membranes, followed by lysis of the bacteria and immobilization of the total protein complement [17, 18]. However, these early attempts only had limited success, because 1) heterologous proteins (e.g., human proteins) were expressed in bacteria, yielding proteins that lacked critical eukaryotic posttranslational modifications; 2) denaturing conditions were used to lyse the bacterial host, resulting in improperly folded proteins; 3) proteins of interest were not purified away from thousands of unwanted bacterial proteins; and 4) the density of the array was low. Before long, other research groups began to report their efforts to fabricate high-density protein

microarrays with purified proteins or antibodies [19-22]. In order to improve protein stability and preserve the native conformation of purified proteins, many research groups developed a variety of surface features to keep proteins hydrated during protein microarray fabrication. These efforts included reports on 3D gel-pad chips [23], nanowell chips [21], and plasma membrane-coated chips [24], to name a few.

The real breakthrough was a 2001 report on the fabrication of a yeast proteome microarray by the Snyder group [25]. In this study, approximately 5,800 full-length yeast ORFs were individually expressed in yeast and their protein products purified as N-terminal GST-fusion proteins. Then, each purified protein was robotically spotted on a single glass slide in duplicate at high-density to form the first "proteome" microarray, as it covered more than 75% of the yeast proteome. More recently, proteome microarrays have been fabricated from the proteomes of viruses, bacteria, plants, and humans [2-5, 8, 10, 26].

On the basis of their applications, protein microarrays can be divided into two classes: analytical and functional protein microarrays [27]. Unlike antibody arrays (analytical microarrays), functional protein microarrays are made by spotting purified proteins on solid surfaces and are therefore useful for direct characterization of protein functions, such as protein binding properties, posttranslational modifications, enzyme-substrate relationships, and immune responses [28]. More recently, a reverse-phase array was developed in which tissue or cell lysates, as opposed to antibodies, are used to construct the array [29].

Meanwhile, we and others have developed various types of biochemical assays that can be conducted using protein microarrays to characterize protein-binding properties, including protein-protein, -DNA, -RNA, and, -lipid interactions, and to identify substrates of various types of enzymes, such as protein kinases, acetyltransferases, and ubiquitin and SUMO E3 ligases via covalent reactions [1, 19, 30-37] (Table 1). These efforts clearly demonstrate the versatility and power of protein microarray technology as a systems biology and proteomics tool. In this review, we will summarize recent applications of protein microarrays in clinical proteomics, including biomarker identification, pathogen-host interactions, and cancer biology (Table 2).

1.2. Biomarker identification

One of the most rapidly growing applications of protein microarray technology in the field of clinical proteomics is biomarker identification. This application for protein microarrays stemmed from traditional serology studies, which focus on the diagnostic identification of antibodies in patient serum samples. These antibodies can be produced as part of an immune response to an infection, against a foreign protein, or even against a person's own proteins. When proteins on a protein microarray are viewed as potential antigens, researchers can use it as a platform to identify autoantibodies that show statistically significant association with an infection or with a disease of interest. In general, the following approach is used: first, patient sera are diluted (e.g., 1000-fold) and incubated on a pre-blocked antigen microarray (i.e., protein microarray), followed by a stringent washing step. Then, positive signals are detected using anti-human IgG, IgM, or IgA antibodies

Table 1. **Protein Microarray Studies by Posttranslational Modification**

PTM	Substrate	Enzyme	Reference
Phosphorylation	Yeast	87 yeast kinases	[1, 32]
	Human	Human CDK5, CKII	[39, 40]
		Four herpesvirus kinases	[41]
	Arabidopsis	*Arabidopsis* MAPKs	[4, 26]
	Herpesvirus	EBV BGLF4	[42]
Ubiquitylation	Yeast	HECT-domain E3 Rsp5	[34, 43]
	Human	HECT-domain E3 Nedd4 & Nedd4L	[44]
Acetylation	Yeast	NuA4 complex	[33, 35]
	E. coli	PAT	[45]
SUMOylation	Human	SAE1/SAE2, Ubc9	[36]
S-nitrosylation	Yeast & Human	N/A	[46]

coupled with various fluorophores for detection (Figure 1). Compared with traditional serology techniques, such as ELISA, agglutination, precipitation, complement-fixation, and fluorescent antibodies, protein microarray-based serum profiling is much more sensitive and can be performed at a much higher throughput. Another significant advantage is that it offers an unbiased platform for novel biomarker identification. In this section, we will review four studies to illustrate the history and development of protein microarrays in biomarker identification.

SARS-CoV diagnosis

In 2003, Zhu et al fabricated the first viral proteome microarray composed of every

Table 2. **Protein Microarray Studies in Clinical Proteomics**

Disease Type	Disease	Detection methods	Ref
Infectious	SARS infection	Serum antibodies detected on SARS-CoV arrays comprised of ~60 purified proteins of 5 coronaviruses	[2]
	B cell lymphoma or AIDS-related Kaposi's lymphoma	Serum antibodies detected on herpesvirus array comprised of ~80 purified EBV and KSHV proteins	[47]
	Rabbit model of Plague	Rabbit serum antibodies detected on *Yersinia pestis* arrays comprised of 149 proteins	[48]
	Brucellosis	Serum antibodies detected on *Brucella melitensis* arrays comprised of 3046 proteins expressed in lysates	[49]
	Cervical carcinomas or precursor lesions	Serum antibodies detected on papillomavirus arrays comprised of 154 proteins of 13 viruses	[50]
	Streptococcus infection	Human proteins detected on Streptococcal surface protein arrays comprised of 201 purified proteins from 2 pathogenic strains	[7]
Autoimmune	Inflammatory Bowel Disease (CD and UC)	Serum antibodies detected on *E. coli* K12 arrays comprised of purified 4,179 proteins	[51]
	Autoimmune Hepatitis	Serum autoantibodies detected on human arrays comprised of 5011 purified proteins	[52]
	Primary Biliary Cirrhosis	Serum autoantibodies detected on human arrays comprised of ~17,000 purified proteins	[53]
	Sjögren's syndrome	Saliva autoantibody detected on human arrays comprised of ~8,000 purified proteins	[54]
Cancer	Breast Cancer	Serum autoantibodies detected on human arrays comprised of 4988 candidate tumor antigens	[55]
		Cancer stem-like cell glycan signature identified using array of 94 lectins	[56]
	Bladder Cancer	Serum autoantibodies detected on human arrays comprised of ~8,000 purified proteins	[57]
	Rhabdomyosarcoma	Phosphorylation status of 27 proteins detected on human arrays spotted with cancer cell lysates	[58]

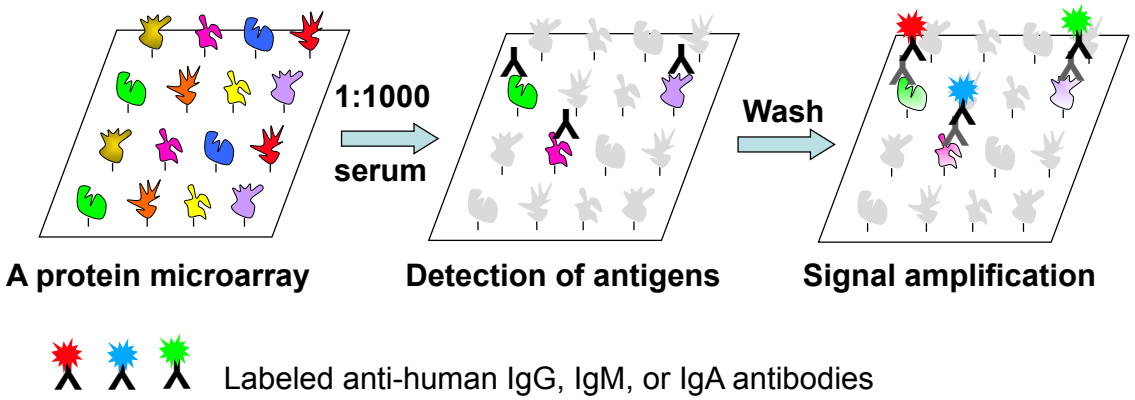

Figure 1. **Principle of serum profiling assays performed on a functional protein microarray.** A functional protein microarray, composed of hundreds of thousands of individually purified proteins, is first blocked with BSA in PBS buffer. Then, a diluted serum sample is incubated on the microarray typically at RT for 1 hr. After extensive washes, bound antibodies (e.g., human IgG, IgA, or IgM) can be detected with anti-human immunoglobulin antibodies, followed by a signal amplification step with fluorescently labeled secondary antibodies. Detection of immunoglobulin isotypes can be multiplexed with different fluorophores as illustrated.

full-length protein and protein fragment encoded by SARS coronavirus (SARS-CoV), as well as proteins from five additional mammalian coronaviruses [2]. These microarrays were then used to screen 400 Canadian serum samples collected during the 2002 SARS outbreak, including samples from confirmed SARS-CoV cases, respiratory illness patients, and healthcare professionals. Antibody response was quantified by the application of both anti-human IgG and IgM antibodies each coupled to different fluorophores, followed by measurement of fluorescence signal intensity (Figure 2). To identify potential biomarkers,

serum samples were first clustered according to the relative signal intensities of all of the coronavirus proteins in an unsupervised fashion (See Data Analysis section). The serum samples fell into two major groups, which upon subsequent comparison with clinical data were largely correlated with either SARS-positive or SARS-negative sera. In the cluster of markers, five fragments of the SARS N protein associated tightly with SARS infection, while SARS sera also exhibited statistically significant binding to one spike protein fragment. However, a few proteins encoded by other coronaviruses also showed significant correlation. To determine the best classifiers and classification model, two different supervised analysis approaches, k nearest neighbor (k-NN) and logistics regression (LR) were applied, and the N protein of SARS CoV, as well as the spike protein S from both SARS CoV and HCoV-229E, were identified as the best classifiers. A useful feature of a serum test relative to a nucleic acid diagnostic test is that anti-pathogen antibodies can potentially be detected long after infection. Taking advantage of this, serum samples collected from SARS patients, who recovered from respiratory disease (~320 days after diagnosis), were used to probe the microarray and positive signals were detected with both anti-human IgG and IgM antibodies (Figure 2; middle panel). These results clearly showed that SARS CoV N proteins could be readily recognized by human IgG antibodies and importantly, not by IgM antibodies, as expected. However, serum samples collected from the Chinese patients immediately after fever was detected showed much stronger signals both in the IgG and IgM profiling (Figure 2; left panel). These results indicated that the protein microarray approach is capable of detecting anti-pathogen antibodies in serum samples long after infection, as well as detecting infection at early stages of infection as demonstrated by anti-human IgM profiling. The

approach developed here is potentially applicable to all viruses and expected to have a great impact on epidemiological studies and possibly in clinical diagnoses.

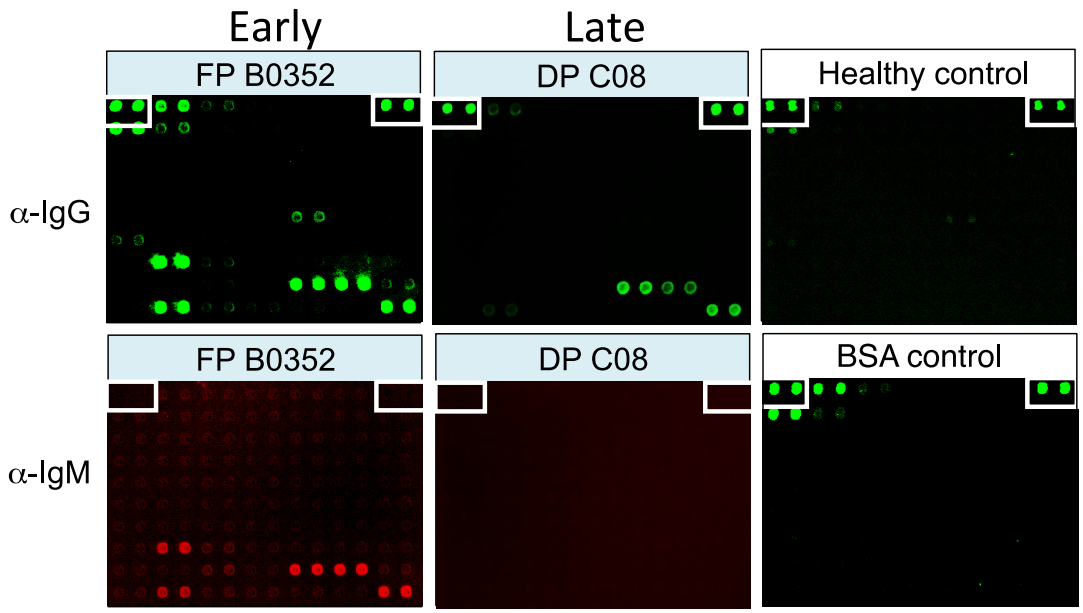

Figure 2. Examples of IgG and IgM profiles obtained with serum samples of SARS-CoV-infected patients. Sample FP B0352 was collected immediately following detection of fever in a patient in Beijing; Sample DP C08 was collected from a recovered SARS patient in Toronto. Signals in the upper panel and the BSA control were detected with anti-human IgGs, while signals in the lower panel of the two patient samples were detected with anti-human IgMs.

Humoral immune responses to herpesviruses

A similar approach has been applied to profile humoral immune responses to two human herpesviruses, Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus

(KSHV). EBV is a ubiquitous human herpesvirus, while KSHV has a restricted seroprevalence. Both viruses are associated with malignancies and show an increased frequency in individuals who are co-infected with human immunodeficiency virus type 1 (HIV-1). The Zhu and Hayward groups generated a protein microarray consisting of 174 EBV and KSHV full-length proteins that were individually expressed and purified from yeast [42, 47]. Instead of sera, plasma antibody responses to EBV and KSHV were examined from healthy volunteers and patients with B cell lymphoma or with AIDS-related Kaposi's sarcoma or lymphoma. These experiments detected IgG responses to known antigens, as well as the tegument proteins ORF38 (KSHV), BBRF (EBV), BGLF2 (EBV), and BNRF1 (EBV), and to the EBV early lytic proteins BRRF1 and BORF2. Because IgA responses to EBV EBNA1 and viral capsid antigens have long been used as a diagnostic tool for nasopharyngeal carcinoma, they also found IgA responses in healthy and HIV-infected patients. IgA responses to VCA and to EBNA1 were found to be frequently elevated in lymphoma patients and in individuals who were HIV-1 positive. Comparison between the IgG and IgA responses indicated that IgA responses were much higher against BCRF1, BRRF2, and LMP2A. Therefore, this study demonstrated that plasma can be used for biomarker identification; immunoglobulin responses of other isotypes, such as IgA, are therefore also worth testing.

E. coli proteome microarrays for IBD diagnosis

To demonstrate that protein microarrays could also be used to identify new biomarkers in autoimmune diseases, Chen et al decided to apply an *E. coli* K12 proteome

microarray [3] to profile serum samples collected from Crohn's disease (CD) and ulcerative colitis (UC) patients [51]. CD and UC are chronic, idiopathic, and clinically heterogeneous intestinal disorders collectively known as inflammatory bowel disease (IBD). Although IBDs have been suggested to be autoimmune diseases, anti-microbial antibodies are present in the sera of IBD patients, and some of these antigens have proven to be valuable serological biomarkers for diagnosis and/or prognosis of the disease. In this study, a protein microarray, including 4,256 proteins encoded by a commensal K12 strain, was screened using individual serum from healthy controls (n = 39) and clinically well-characterized patients with IBD (66 CD and 29 UC). Surprisingly, among the 417 *E. coli* proteins that were differentially recognized by serum antibodies from healthy controls and either CD or UC patients, 169 proteins were identified as highly immunogenic in healthy controls, 186 proteins were identified as highly immunogenic in CD patients, and only 19 proteins were identified as highly immunogenic in UC patients. Using several statistical tools, they identified two sets of serum antibodies as novel biomarkers for specifically distinguishing CD from healthy controls (accuracy, 86±4%; $p < 0.01$) and CD from UC (accuracy, 80±2%; $p < 0.01$). This study was the first demonstration of using high-density, high-content proteome microarrays to discover novel serological biomarkers. It was also the first effort to examine human immune responses to the entire proteome of a microbial species in a disease context.

Autoantigen discovery for AIH

A protein microarray composed of individually purified human proteins would be an ideal tool for discovery of novel autoantigens associated with an autoimmune disease. Take

autoimmune hepatitis (AIH) as an example: AIH is a chronic necroinflammatory disease of human liver with little known etiology. Detection of non-organ-specific and liver-related autoantibodies using immunoserological approaches has been widely used for diagnosis and prognosis. However, these traditional autoantigens, such as anti-SMA (smooth muscle autoantibodies) and anti-ANA (antinuclear autoantibodies) are often mixtures of complex biological materials. Unambiguous and accurate detection of the disease demands identification and characterization of these autoantigens. Therefore, Song et al fabricated a human protein microarray of 5,011 non-redundant proteins that were expressed and purified as GST fusions in yeast [36]. There are several advantages associated with producing human proteins in yeast rather than bacteria: 1) higher solubility, 2) higher yields of large proteins (e.g. > 50 kD), 3) better preserved conformation of proteins, and 4) proteins are less immunogenic when produced in yeast than in *E. coli* [3, 10, 31]. However, unlike a viral or bacterial protein microarray, a significant obstacle to the use of a human protein microarray of high content is its high cost. For example, a human protein array of 9,000 proteins can exceed $1000 per array. In order to reduce this cost, Song et al developed a two-phase strategy to identify new biomarkers in AIH. Phase I is designed for rapid selection of candidate biomarkers, which are then validated in Phase II (Figure 3). In Phase I, 30 AIH and 30 control serum samples were selected and individually used to probe the human protein microarrays at a 1000-fold dilution, followed by detection of bound human autoantibodies using a Cy-5-conjugated anti-human IgG antibody. Statistical analysis revealed 11 candidate autoantigens. To validate these candidates and to avoid a potential overfitting problem (see below), which is especially likely when dealing with a small sample size, the 11 proteins and 3 positive controls were re-purified to build a large number of low-cost small arrays for

Phase II validation. These arrays were then sequentially probed with serum samples used in Phase I and serum samples obtained from an additional 22 AIH, 50 primary biliary cirrhosis (PBC), 43 hepatitis B (HB), 41 hepatitis C (HC), 11 system lupus erythematosus (SLE), 11 primary Sjögren's syndrome (pSS), and 2 rheumatoid arthritis (RA) patients. As negative controls, they also included 26 serum samples from patients suffering from other types of severe disease and 50 samples from healthy subjects. Three new antigens, RPS20, Alb2-like, and dUTPase, were identified as highly AIH-specific biomarkers with sensitivities of 47.5% (RPS20), 45.5% (Alba-like), and 22.7% (dUTPase), which were further validated with additional AIH samples in a double-blind design. Finally, they demonstrated that these new biomarkers could be readily applied to ELISA-based assays for clinical diagnosis and prognosis.

This study represents a new paradigm in biomarker identification using protein microarrays for three reasons. First, a manageable number of candidate biomarkers can be rapidly identified at low cost because fewer expensive protein microarrays of high-content are needed in the first phase of this two-phase strategy. Second, by using small arrays comprised of selected candidate proteins, the validation step can be rapidly carried out with a much larger cohort at lower cost. This validation step is extremely important for avoiding the overfitting problem associated with statistical analysis in biomarker or classifier identification, especially when dealing with a small cohort (e.g., <40).

In addition, there have seen a series of studies that employed pathogen protein microarrays to profile serological responses following infection. For example, protein

microarrays have been developed in bacteria and viruses for biomarker identification in various infectious diseases [37-40]. These studies have clearly demonstrated the power of protein microarrays in identification of potential biomarkers; however, several shortcomings are repeatedly seen in these studies. For instance, many of these arrays were fabricated using proteins translated in *E. coli* lysates without purification [49, 52, 59]. Because these proteins are contaminated with unwanted *E. coli* proteins, sensitivity of the assay is likely reduced due to their high immunogenicity [47]. As a result, *E. coli* lysates had to be used as a blocking reagent to alleviate this problem. Also problematic is that in many of these studies, identified biomarkers were not validated with additional cohorts and therefore, the possibility of overfitting was not completely ruled out.

Figure 3. Scheme of the two-phase autoimmune disease biomarker strategy. In Phase I, a small cohort is used to rapidly identify a group of candidate biomarkers via serum profiling assays on a human protein microarray of high cost. Because a small number of microarrays are needed, cost of the experiments is relatively low. In Phase II, a focused protein microarray of low cost is fabricated by spotting down purified candidate proteins. A much larger cohort is then assayed on these arrays in a double blind fashion to validate the candidates identified in Phase I.

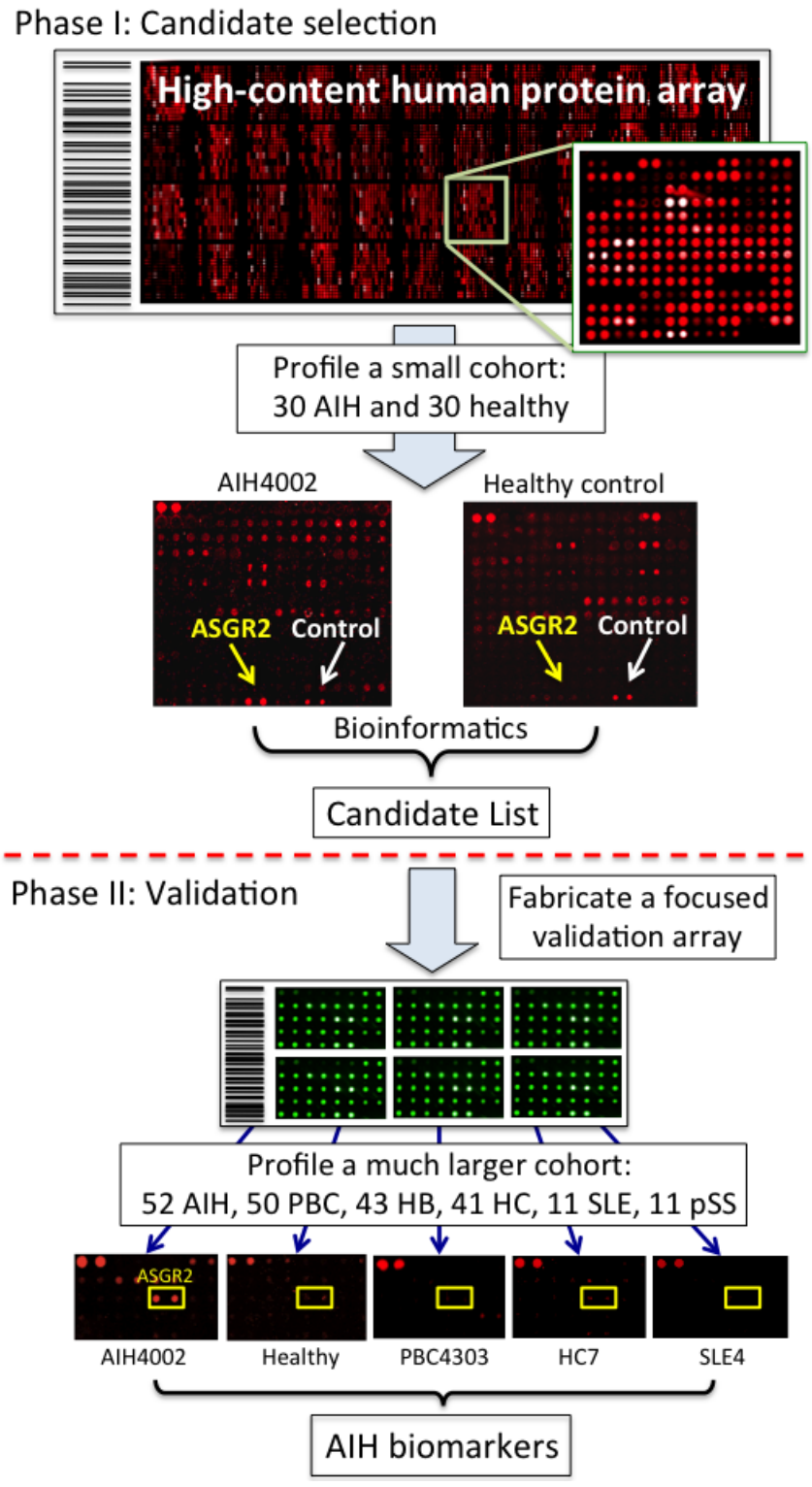

Figure 3. Scheme of the two-phase autoimmune disease biomarker strategy.

1.3. Pathogen-host interactions

An emerging application of protein microarrays in the field of clinical proteomics is an unbiased, proteome-wide survey of important players involved in pathogen-host interactions. The identified factors, encoded by either a pathogen or a host, have the potential to be developed into novel therapeutic targets. Protein microarrays can serve as an ideal platform for such purposes: Once a protein microarray is fabricated from a host or pathogen, it can be used to identify direct pathogen-host interactions. This strategy is particular useful for investigating virus-host interactions because after entering the host cells, the viral genome and encoded proteins are in direct physical contact with the host's biological materials. As we will discuss in this section, such interactions can be investigated at multiple levels, such as RNA-protein interactions, enzyme-substrate relationships, and protein-protein interactions.

BMV RNA and host proteome interactions

In 2007, Zhu et al described the first study using a yeast proteome microarray to identify host factors that can affect replication of Brome Mosaic Virus (BMV), a plant-infecting RNA virus that can also replicate in *S. cerevisiae* [31]. Previous studies have shown that this positive-stranded RNA virus encodes a tRNA-like structure at the 3'-end of its RNA genome, in which a clamped adenine motif (CAM) is required for packaging its genome into the capsid. To identify crucial host proteins that can interfere with the viral packaging process, a Cy3-labeled CAM-containing RNA stem-loop structure was incubated on the yeast proteome microarray in the presence of an equal amount of a Cy5-labeled

mutated CAM hairpin. Using Cy3-to-Cy5 fluorescence signal intensity ratios, the top hits were identified and validated using an in vitro gel-shift assay. Two validated candidate proteins, pseudouridine synthase 4 (Pus4) and actin patch protein 1 (App1), were selected for further characterization in tobacco plants. Both proteins modestly reduced BMV genomic plus-strand RNA accumulation, but dramatically inhibited BMV systemic spread in plants. Pus4 also prevented the encapsidation of the BMV RNAs in plants and the reassembly of BMV virions in vitro.

This work is significant because it established the first RNA-binding assay on a proteome microarray and demonstrated the utility of protein microarrays for identifying important players involved in pathogen-host interactions.

Host phosphorylome of virus-encoded kinases

In the course of evolution, viruses have been very successful at exploiting the host via development of their own arsenals, some of which were hijacked from the host in the form of both DNA and proteins. To develop more effective antivirals, one must understand the molecular mechanisms by which viruses exploit the host machineries for their own use. The human α, β, and γ herpesviruses infect different tissues and cause distinct diseases, ranging from mild cold sores to pneumonitis, birth defects, and cancers [60]. However, they each confront many of the same challenges in infecting their hosts, including reprogramming cellular gene expression, sensing cell-cycle phase and modifying cell-cycle progression, and reactivating the lytic life cycle to produce new virions and spread infection. On the other

hand, many lytic cycle genes involved in replication of the viral genomes (e.g., the orthologous serine/threonine protein kinases) are highly conserved across the herpesvirus family. Therefore, it became an attractive hypothesis that the shared substrates targeted by these orthologous viral kinases would reveal host pathways that are critical for replication across the herpesvirus familyTo test the above hypothesis, Li et al employed a human protein microarray [42]. The authors purified four orthologous kinases encoded by EBV, KSHV, HCMV, and HSV-1, performed kinase reactions on a human protein microarray described previously [4], and identified 110 shared substrates. Like every large-scale screen, the next challenge was to select candidates that would be worth pursuing. To do so, the authors then applied Gene Ontology (GO) and STRING analyses (http://string-db.org/, a database of known and predicted protein-protein interactions) to these candidates and found a highly connected cluster of 15 proteins. Strikingly, these proteins were all known to be involved in the DNA damage response (DDR) (Figure 4). The host DDR has been known to be important to many viruses, including human herpesviruses, and is relevant to virus-induced tumorigenesis [41]. To narrow down this list to a single candidate for in-depth characterization, the authors reasoned that the viruses are likely to target an upstream master regulator, which triggers the DNA damage response. On the basis of a literature search and the structure of this cluster, Tat-interactive protein 60 (TIP60) emerged as an excellent candidate for follow-up, because 1) TIP60 is further upstream in the DDR pathway than any of the other candidates in the cluster; 2) it serves as a master regulator in DDR via activation of ATM autophosphorylation activity by acetylation; 3) it regulates chromatin dynamics via histone acetylation; and 4) its importance has been shown in other viruses. Indeed, the authors observed that when TIP60 was knocked down in EBV-infected replication was

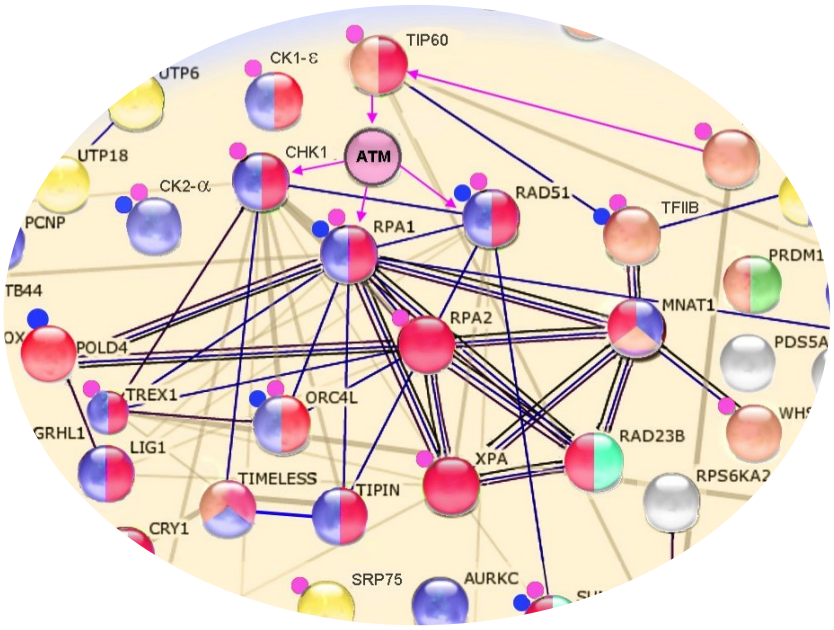

Figure 4. Identification of most relevant candidate for in-depth in vivo studies. Assisted by GO analysis, 110 shared substrates of conserved herpesvirus kinases were plugged into the STRING database. A highly connected cluster of 15 proteins was revealed, all of which are known to play a role in DDR. Based on the literature and topology of the cluster, TIP60 emerged as the most promising candidate. Protein nodes are color coded by functional class, with proteins involved in DDR colored red. Small blue circles adjacent to protein nodes indicate that the protein is herpesvirus associated, while small pink circles indicate the proteins are associated with other viruses. Edges between the proteins represent known or predicted connections, such as protein-protein interactions, catalytic reactions, and enzyme-substrate relationships.

greatly reduced. Next, the authors applied a series of cell-based assays and showed that during EBV replication, TIP60 activation by the BGLF4 kinase triggers EBV-induced DDR and also mediates induction of viral lytic gene expression. Finally, the authors demonstrated that TIP60 was also required for efficient lytic replication in HCMV, KSHV, and HSV-1.

This work illustrates the value of high-throughput, unbiased approaches for the discovery of conserved viral targets in the host that have the potential to be developed into novel therapeutic targets for antivirals. Currently, there are few drugs available to treat herpesvirus infections, and viral escape mutants develop as a result of extensive use of this limited repertoire. The herpesvirus protein kinases are attractive antiviral drug targets. However, developing broadly effectively drugs targeting protein kinases requires knowledge of their common cellular substrates. The information provided by common substrate identification will assist in the design of assays for new and broadly effective anti-herpesvirus therapeutics.

LANA interactome analysis reveals a role in telomere shortening

Protein microarrays can also serve as a convenient tool for profiling protein-protein interactomes between a pathogen and a host. In a recent example, Hayward and colleagues surveyed the interactome between a KSHV-encoded virulent factor, LANA, and the human host using human protein microarrays, in order to identify host proteins that can be recognized by LANA [44]. LANA functions in latently infected cells as an essential participant in KSHV genome replication and as a driver of dysregulated cell growth.

Although yeast two-hybrid screens, glutathione *S*-transferase (GST) affinity, immunoprecipitation (IP) assays, and chromatography coupled with mass spectroscopy have been applied to the identification of LANA binding proteins, each approach has strengths and weaknesses, and each tend to identify different sets of proteins. In this study, the authors used purified FLAG-tagged LANA applied to human protein microarrays to identify 61 potential binding partners, many of which were previously unknown. 8 out of 9 proteins were validated by co-immunoprecipitation, including TIP60, protein phosphatase 2A (PP2A), replication protein A (RPA) and XPA. Although human papillomavirus (HPV) E6, HIV-1 TAT, and human cytomegalovirus (HCMV) pUL27 interact with TIP60 and induce TIP60 degradation, LANA-associated 42 retained acetyltransferase activity and showed increased stability. This observation is in line with the study described in the previous section that showed that TIP60 plays a positive role in KSHV lytic replication. On the other hand, identification of RPA as a LANA interacting partner suggested that LANA may play a role in regulating the length of host telomeres, because RPA1 and RPA2 are known to be essential in the replication of cellular telomeric DNA. To test this hypothesis, the authors performed ChIP assays with anti-RPA1 and -RPA2 antibodies using primers specific to the telomere regions and found that the presence of LANA drastically reduced the recruitment of both RPA1 and RPA2 to the host telomeres, while it had no impact on the protein level of the RPA complex. This observation raised the possibility that LANA might have an impact on telomere length. Using Southern blot analysis of terminal restriction fragments, the standard method for quantifying telomere length, the authors demonstrated that the average length of telomeres was shortened by at least 50% in both LANA-expressing endothelial cells and KSHV-infected primary effusion lymphoma cells. Many interesting questions remain to be

answered. How does LANA block the RPA complex recruitment to the telomeres? Is it achieved via direct competition since LANA is also a ssDNA binding protein? Or, does LANA serve as a kinase sink for the RPA complex and regulate RPA recruitment via phosphorylation?

SUMO-EBV interactome revealed a new mechanism of EBV lytic replication

On the flip side, a human factor of interest can be used to survey a virus protein microarray to identify important viral factors. Similar to the ubiquitylation pathway, SUMOylation involves a series of sequential enzymatic reactions that conjugate SUMO to lysine residues on substrate proteins. Previous studies have shown that both latent and lytic EBV proteins interact with the SUMO system. Noncovalent SUMO-EBV protein interactions can occur via a SUMO interaction motif (SIM) in the target proteins. To comprehensively identify additional EBV proteins that bind to SUMO, Li et al performed a protein-binding assay with human SUMO2 [61] on a previously described EBV proteome microarray [47] and identified a total of 11 proteins, including the conserved viral kinase BGLF4. The mutation of potential SIMs in BGLF4 at both N- and C-termini changed the intracellular localization of BGLF4 from nuclear to cytoplasmic, while BGLF4 mutated in the N-terminal SIM remained predominantly nuclear. The mutation of the C-terminal SIM yielded an intermediate phenotype with nuclear and cytoplasmic staining. The authors also found that BGLF4 abolished the SUMOylation of the EBV lytic cycle transactivator ZTA, and that this inhibitory effect on ZTA SUMOylation was dependent on both BGLF4 SUMO binding and BGLF4 kinase activity. The global profile of protein SUMOylation was also suppressed by

BGLF4 but not by the SIM or kinase-dead BGLF4 mutant. Furthermore, BGLF's interaction with SUMO was required to induce the cellular DNA damage response and to enhance the production of extracellular virus during EBV lytic replication.

Identification of novel streptococcal proteins that bind human ligands

The identification of pathogen proteins that interact with human factors has also been applied to understanding the mechanisms of bacterial infection. Margarit and others harnessed the power of protein microarrays to identify proteins expressed by two species of the streptococcus gram-positive bacteria, *Streptococcus pyogenes* and *S. agalactiae*, that interact with human factors known to mediate pathogenesis [62]. Rather than develop whole-proteome arrays, they used a bioinformatics approach to predict those proteins present on the cellular surface—and thus most likely to play a role in infection, and used this list of 200 proteins to develop their arrays. They also carefully considered the human probes that they would use, choosing three human ligands: fibronectin, fibrinogen, and C4 binding protein, all known to play important roles in the colonization and infection processes. Binding experiments conducted using the streptococcal arrays and human protein probes identified 17 of the 20 known interactions previously reported as well as 8 newly identified streptococcal proteins, many of which they confirmed by far-western blot analysis. These novel proteins included proteins of unknown function as well as 3 related proteins that they termed the fib proteins. They then used domain mapping to identify regions of the fib proteins required for their interaction with the human ligands. Interestingly, sera samples from patients with *S. agalactiae* infections show high titers of Fib-specific antibodies, indicating that these proteins are highly expressed during infection. Further work will determine the role of these

proteins in infection and whether they will emerge as suitable drug targets to fight pathogenic Streptoccus infections.

In summary, the above studies have demonstrated the power of protein microarrays in the discovery of novel molecular mechanisms underlying host-pathogen interactions at various levels. In recent years, other high-throughput approaches, such as shotgun mass spectrometry [63], genome-wide RNAi screens [64, 65], and yeast two-hybrid [66, 67] have been applied to understand host-pathogen interactions; however, the protein microarray approach provides a more versatile platform than any of these single approaches for identifying multiple types of direct interactions between a pathogen and host, including protein-protein interactions [61, 62, 68], RNA-protein interactions [31], and enzyme-substrate interactions [41].

1.4. Cancer biology

Over the past five years, rapid development of genome-wide sequencing technologies (i.e., next-gen sequencing) has revealed the heterogeneous nature of tumors [69, 70]. However, clinical diagnosis of tumors is largely still dependent on morphologic patterns. The fact that tumors with indistinguishable morphology can have vastly different clinical outcomes suggests that the molecular heterogeneity of each patient's tumor cells have to be better understood before more effective therapies can be developed. Therefore, the future of cancer treatment is tailored molecular therapy specific for each individual, which will require a new class of proteomic profiling technologies. As a widely adopted technology, protein

microarrays can meet this need for the profiling of the functional state of tumors and for cancer biomarker identification.

Identification of Autoantibody Biomarkers for the early detection of breast cancer

Current screening for breast cancer using mammograms detects only 70% of breast cancers, and false-positive mammograms lead to unnecessary biopsies. The identification of biomarkers that would allow early detection of breast cancer could provide a non-invasive, low cost method that could improve patient outcomes. One promising category of cancer biomarkers are autoantibodies to tumor antigens which offer better stability, specificity, ease of purification, and ease of detection compared to other serum proteins. In order to identify autoantibodies to tumor antigens associated with breast cancer, Anderson et al. used protein arrays containing candidate tumor antigens and applied breast cancer patient and control serum samples to identify differences in the human antibody repertoire that could be used as biomarkers [55]. These custom protein arrays, termed "NAPPA" arrays (Nucleic Acid Protein Programmable Array), are fabricated by the spotting of cDNAs that encode the target proteins at each feature of the array. Proteins are transcribed and translated by a cell-free system and immobilized by encoded epitope tags, thus bypassing the protein purification process. Additionally, the authors used a three-phase screening approach to home in on the best candidate breast cancer biomarkers. In the first phase, they used arrays with the full set of 4988 tumor antigens in order to eliminate uninformative autoantibodies that were present at similar levels in both early breast cancer patients and healthy women. Subtracting these antigens, the protein set was reduced to 761, allowing them to fabricate smaller arrays for the

next phase that offer the benefits of reduced cost and fewer false positives. In the second phase, sera from patients with invasive early breast cancer and benign breast disease were compared, in order to identify antigens specific to early breast cancer but absent from benign breast disease, resulting in 119 antigens. In the third phase, they set out to validate this antigen list, finding 28 antigens that maintained high levels of specificity in a blinded validation assay, including the protein ATP6AP1, a known autoantigen. They then focused on this protein and went on to show high expression of ATP6AP1 in 4 breast cancer cell lines by Western blot, as well as significantly higher ATP6AP1 autoantibody levels in ~13% of early breast cancer serum cases compared to controls. Although only a first step, this work demonstrates the power of protein microarrays, in particular programmable protein microarrays, in identifying biomarkers for the early detection of breast cancer.

Finding autoantibody biomarkers in bladder cancer

An important goal of identifying cancer biomarkers is to define new strategies for early diagnosis that can allow early intervention with current therapies to improve patient survival rates. Additionally, since cancer-associated autoantibodies often target proteins that are mutated, modified, or aberrantly expressed in tumor cells, they could also be considered immunologic reporters that could uncover molecular events underlying tumorigenesis [57]. The molecular players in these events, in turn, may be the best place to start in efforts to develop novel therapies. In order to identify autoantibody biomarkers that could act as indicators of bladder cancer, as well as the underlying molecular pathology contributing to disease, Orenes-Pinero turned to a protein array strategy using the Invitrogen Protoarray

containing ~8,000 purified human proteins to identify antibodies to tumor-associated antigens in serum. Comparing serum samples collected from 12 patients with bladder cancer and 10 control patients without bladder cancer, they identified 171 differentially expressed proteins. Among these, they selected clusterin and dynamin for validation based in part on their known role in cancer biology. Using immunohistochemistry on a custom tissue microarray comprised of bladder cancer tumor samples, they found reduced expression levels of clusterin in muscle invasive bladder tumors as compared to nonmuscle invasive tumors. On the other hand, they found that low protein expression of dynamin was associated with increased tumor stage and grade, higher recurrence rate after surgery, as well as shorter survival. Paradoxically, their follow-up tests revealed lower expression levels of dynamin and clusterin associated with disease, in contrast to their protein array results which showed increased autoantibody levels to these proteins among bladder cancer patients compared to controls. Despite these contradictory findings, the authors demonstrated significant associations between dynamin and clusterin expression levels and bladder cancer disease progression that could potentially allow them to use these as informative biomarkers in the clinic as well as potential drug targets. This work demonstrates the power of protein microarrays for the identification of autoantibodies to tumor-associated antigens and its application to the discovery of cancer biomarkers.

1.5. Outlook

Recent years have witnessed tremendous growth in the use of protein microarrays to address important questions in the field of clinical proteomics. In the area of biomarker

identification, most of the recent research has been focused on either infections or autoimmune diseases. We believe that protein microarrays, especially functional protein microarrays, will be widely used for identification of cancer biomarkers in the near future. Indeed, recent advances in immunoproteomics and high-throughput technologies have suggested that the autoantibody repertoire in cancer patients might be quite different as compared with that in healthy subjects, leading to the hypothesis that autoantigens might be identified as biomarkers for cancer diagnosis, as well as cancer prognosis [71]. Ideally, a human protein microarray developed for such a purpose should cover the entire human proteome, in order to enable a comprehensive screen for the autoantigens. To our knowledge, we have fabricated a human proteome microarray of the best coverage (>70%) [10]. However, when hundreds, if not thousands, of serum samples are needed to screen for biomarkers, the cost of using these human proteome microarrays accumulates very rapidly. An effective strategy to overcome this obstacle is to apply the two-phase strategy as described in the AIH study [7]. We expect that this strategy will become popular in the near future. Finally, we expect that functional protein microarrays will be used as a readout to obtain reaction profiles of the collected activities of various types of enzymes, such as kinases, acetyltransferases, ubiquitin and SUMO E3 ligases in cancerous tissues. Comparing PTM profiles obtained from cancer and healthy tissues will allow us to identify biomarkers and to gain new insights into the molecular mechanisms of disease.

Chapter 2

Introduction to SUMOylation

2.1. The SUMOylation pathway

Posttranslational modifications of proteins provide a central molecular mechanism for cells to respond to external stimuli [72]. While phosphorylation is perhaps the best-characterized posttranslational modification, numerous other modifications play crucial roles in cellular responses to their environment. The UBL (ubiquitin-like) family of posttranslational modifications is distinct in that it involves attachment of a protein to target proteins. Many types of UBLs exist, including ubiquitination, SUMOylation, ISGylation, and NEDDylation, URMylation, ATGylation, and FUBIlation. Protein SUMOylation is an essential process in most organisms, including S. cerevisiae, C. elegans, Arabidopsis thaliana and mice [73]. SUMOylation is the best characterized of the UBL-modifications other than ubiquitin itself and has been implicated in regulating a diverse array of protein activities including localization, inhibition of transcription factors, inhibition or activation of enzymes, as well as stimulating, or preventing, degradation [74]. Despite a large body of literature demonstrating a wide range of functions for SUMOylation, many important questions remain about its physiological roles in the cell.

SUMOylation was first discovered almost 20 years ago in 6 different labs independently. The first publication reporting the existence of SUMO described a genetic approach that identified the yeast homolog of SUMO, SMT3, as a genetic suppressor of Mif2, a gene involved in mitosis [75]. A year later, another group used a biochemical approach, finding that SUMO was covalently linked to the nuclear protein RANGAP1 in rat liver nuclear envelopes [76]. That same year, work in the laboratory of Edward Yeh identified a

protein that they termed Sentrin as a Tumor necrosis factor death domain-interacting protein in a yeast two-hybrid screen [77]. Meanwhile, another yeast two-hybrid assay identified SUMO, which they termed PIC1, as a protein that interacted with promyelocytic leukemia (PML) protein [78]. In yet a third yeast two-hybrid assay designed to identify proteins that interact with RAD51 and RAD52, two proteins involved in DNA recombination and repair, the protein was found and termed UBL1, or ubiquitin-like 1, due to its homology with ubiquitin [79]. Finally, in 1997, Frauke Melchior and colleagues also described a form of RanGAP1 modified by a novel protein, in rat liver extract, HeLa and other human cell lines, that they named SUMO-1, the name that would later become the official gene symbol [80].

While there is a single SUMO paralog in invertebrates, there are 4 SUMO paralogs in humans. Whether SUMO4 can be conjugated to target proteins is controversial and as a result most SUMO research has focused on the other SUMO paralogs, SUMO1-3. SUMO2 and SUMO3 share 96% amino acid sequence identity and cannot be distinguished by available antibodies. For this reason they are often treated as a single protein, described in the literature as "SUMO2/3". Human SUMO1 shares about 45% sequence identity with SUMO2 and SUMO3. Although structurally very similar, SUMO1 and SUMO2/3 show some functional differences, particularly regarding cellular localization, their role in the formation of SUMO polymers or chains, and to what proteins they bind and are conjugated to. SUMO1 is uniquely found within the nucleoli, nuclear envelope as well as cytoplasmic foci, while SUMO2/3 was found to accumulate on chromosomes [81]. Because SUMO2 contains a SUMOylation consensus site, it is believed that only SUMO2 can act as an acceptor in the

formation of SUMO chains, while SUMO1 is more likely to act as a SUMO chain terminator [82].

A few substrates have been identified that show preferential binding to and/or modification by one or other of the SUMO paralogs. For example, Ran GTPase activating protein 1 (RANGAP1), is preferentially modified by SUMO1 [83]. Ubiquitin specific protease 25 (USP25) and Topoisomerase II are both preferentially modified by SUMO2/3 [84, 85]. The Bloom's syndrome helicase protein (BLM) preferentially binds to SUMO2/3 and this binding is thought to be important for its subsequent SUMO2/3 modification [86].

All UBLs, including SUMOylation, appear to be conjugated to their target proteins by similar enzymatic pathways but generally involve distinct enzymes specific to the particular UBL. The SUMOylation cycle includes 5 major steps: 1) maturation, in which the C-terminus of SUMO is removed by a sentrin protease (SENP) to reveal a reactive diglycine motif, 2) activation, in which SUMO is first adenylated and then transferred to a cysteine on the SUMO E1 heterodimer UBA2/AOS1, 3) conjugation, in which SUMO is transferred to a cysteine on the E2 enzyme UBC9, 4) ligation, in which SUMO is attached to a target protein lysine by UBC9, often in coordination with a SUMO E3 ligase and finally 5) deconjugation, in which a sentrin protease removes SUMO from the target protein.

2.2. Role of SUMO E3 ligases in SUMOylation specificity

Although a SUMOylation consensus motif has been described that frequently provides a site for SUMO attachment to proteins, increasing numbers of proteins have been identified with SUMOylation sites not contained within this consensus motif. The SUMOylation consensus motif sequence, described as ΨKx(D/E), in which Ψ represents a hydrophobic residue, has been shown to bind directly to the SUMO E2 enzyme UBC9, and this direct interaction may explain why the SUMO E1 enzyme and UBC9 alone are sufficient to SUMOylate many substrates in vitro. However, many substrates show enhanced SUMOylation in the presences of a SUMO E3 ligase and it has been proposed that SUMO E3 ligases may be particularly important for SUMOylation of susbtrates in vivo and especially at nonconsensus SUMOylation sites. Furthermore, as the only classes of SUMOylation enzymes for which multiple members have been identified, the SUMO E3 ligases and the SUMO proteases have been proposed to be the major factors determining substrate specificity.

Although there is a lack of consensus in the literature as to the total number of bona fide SUMO E3 ligases, most reported E3s can be divided into three major groups: those that lack any clear consensus motif, those that contain a RING domain, and those that contain a related motif known as an SP-RING domain. SUMO E3 ligases that feature an SP-RING motif may or may not require it for their function, depending on the substrate [87-92]. E3s in this class are characterized by direct binding to their respective targets, the SUMO E2 enzyme UBC9 as well as SUMO itself. The SP-RING motif containing SUMO E3 ligases are

comprised of the human homolog of the yeast protein Mms21, known as NSE2, as well as the PIAS family proteins, PIAS1-4. Included in the category of SUMO E3 ligases that lack any consensus motif is the best characterized SUMO E3 ligase, RANBP2, which is believed to function by stimulating release of SUMO from the SUMO charged-UBC9 thioester complex. SUMO E3 ligases that contain a RING domain include TOPORS and MUL1, two proteins that had been first identified as ubiquitin E3 ligases and were then found to additionally possess SUMO E3 ligase activity [93, 94]. There are currently nearly 617 genes predicted to encode ubiquitin E3 ligases [95]. While there is no clear consensus in the literature on the number of bona fide SUMO E3 ligases, the most liberal estimate would be under 30. Due to the similarities between the ubiquitylation and sumoylation pathways, it is predicted that more SUMO E3 ligases remain to be identified.

2.3. Non-covalent interactions with SUMO

Although SUMOylation can regulate a diverse array of protein activities, at the molecular level SUMOylation alters protein surfaces and their ability to interact with other proteins [73]. It has been proposed that SUMO is a molecular glue that mediates non-covalent interactions between modified substrates and SUMO-binding proteins [96]. This has been demonstrated for the interaction between promyelocytic leukemia (PML) protein and thymine DNA glycosylase (TDG), in which SUMO-modification as well as SUMO-interaction motifs on both proteins promote their interaction. They found that the PML binding activity of SUMO1-modified TDG was 2.5-fold greater than that of the unmodified form in an in vitro assay. They also found that TDG interacted more strongly with the

wildtype form of PML as opposed to the PML(3KR) SUMOylation-defective mutant. These results corroborate the idea that SUMOylation promotes the interaction between TDG and PML.

The transcription factors P300 and Elk-1 each recruit specific HDACs following modification by SUMO1 [97-99]. Another SUMO-mediated protein-protein interaction has been described for RanGAP1 and RANBP2. In this example, SUMO1 modification functions to target RanGAP1 to the nuclear pore complex through stimulating its interaction with RanBP2 [100]. The interaction between the DNA helicase Srs2 and SUMO-modified PCNA recruits Srs2 to replication forks and prevents recombination events during DNA replication [101].

The first SUMO interaction motif (or SIM) was identified through yeast two-hybrid experiments that identified the SUMOylation enzyme SAE2, PML and the PIAS proteins as SUMO interactors [102]. Alignment of the amino acid sequences of these proteins revealed a SUMO interaction motif consisting of an S-X-S motif preceded by hydrophobic amino acids and followed by acidic amino acids. Revisiting this proposed SIM and the importance of individual amino acids that comprise this SIM, subsequent work using NMR found that the S-X-S sequence was less important than the hydrophobic residues and defined a new SUMO interaction motif as V/I-X-V/I-V/I [103]. Later, another yeast two-hybrid focused on the identification of proteins interacting with both SUMO1 and SUMO2 isoforms identified 20 SUMO interacting proteins. Amino acid sequence alignment and subsequent NMR validation supported the importance of a hydrophobic core sequence as well as acidic residues that

could either precede or follow this hydrophobic core, which would determine the orientation of the SUMO and SUMO binding protein interaction [104]. Zinc finger motifs have recently been described as a second type of distinct SUMO interaction motif [105, 106].

2.4. SUMO in human disease

SUMO has been shown to be linked to devastating neurodegenerative diseases including Huntington's and Parkinson's diseases, usually in the context of promoting target protein solubility and preventing the formation of aggregates that are correlated with disease [107, 108]. SUMOylation has also been linked to cardiovascular disease. Recently, it was shown that SUMO gene therapy improved cardiac function in mice with heart failure by promoting the stability of a cardiac calcium ATPase [109]. Finally, the study of SUMOylation is likely to yield important insights into cancer development and therapeutics. High expression levels of a de-SUMOylation enzyme have been correlated with prostate tumor growth and metastasis [110]. SUMO has found to have many roles in different types of DNA damage repair, a process which if better understood could allow the development of more targeted and less toxic chemotherapeutics [111].

2.5. Screens to identify SUMO binding proteins

Takahashi et al used a yeast two-hybrid approach to identify SUMO1 binding proteins [96] and found 5 proteins. Among these included two proteins known to be SUMO-modified, Sp100 and TDG, two proteins known to be involved in SUMO-conjugation, Uba2

and PIAS3, as well as one protein with no previously known interaction with SUMO, EBI3. However, the scope of their study was limited by the use of only the SUMO1 isoform as bait and the use of only 3 types of cDNA libraries, including libraries derived from two immune cell types from mouse, B and T cells, and one library derived from human brain.

The following year, Hecker et al also used a yeast two-hybrid approach to identify SUMO binding proteins, but this time expanded the screen to look at both SUMO1 and SUMO2 paralogs as bait, with cDNA libraries derived from three human tissue types, including thymus, spleen and kidney. They identified 20 proteins including 8 previously known to interact with SUMO and 8 novel SUMO interactors [104]. Among these, most interacted equally with both SUMO1 and SUMO2, except for two proteins. RANBP2, which showed a clear preference for interacting with SUMO1, and TTRAP, which interacted more strongly with SUMO2 than SUMO1.

Ouyang et al used an affinity purification strategy using HeLa nuclear extracts to isolate proteins that interacted with GST-SUMO2 and identify these proteins by mass spectrometry [112]. This approach identified 107 SUMO interacting proteins, representing a large increase in the total number of known SUMO binders. However, this study missed many other potential SUMO interactors including all proteins with subcellular localization outside the nucleus as well as proteins that may bind specifically to SUMO1.

Chapter 3

The INO80 Chromatin Remodeling Complex and Roles of SUMOylation in Chromatin

3.1. The INO80 Chromatin Remodeling Complex

The INO80 chromatin remodeling complex is a 15 subunit complex that has been implicated in the regulation of transcription, checkpoint regulation, chromosome segregation, telomere maintenance, as well as DNA repair and replication [113]. Functional analysis of INO80 chromatin remodeling activity in yeast has primarily focused on its role in histone subunit exchange, while studies of the orthologous mammalian complex have demonstrated a role in nucleosome sliding [114, 115]. The human INO80 complex contains 6 metazoan-specific subunits, including the deubiquitylating enzyme UCH37, TFPT, INO80D, INO80E, MCRS1, and NFRKB, and these subunits are associated with the N-terminal domain of INO80. Although this N-terminal domain is not essential for nucleosome sliding, it has been proposed that these subunits are likely to have regulatory roles in vivo [116].

3.2. The INO80 Complex Subunit TFPT

Although TFPT had been proposed to play a regulatory role as a subunit of the INO80 complex, its function in this context is unknown. Analysis of TFPT function has mostly focused on its role apart from the INO80 complex, and its role in stimulating apoptosis and localization to the nucleus. TFPT was first described in the context of its involvement in a chromosomal aberration underlying pre-B-cell acute lymphoblastic leukemia. This chromosomal aberration resulted in a protein fusion with the transcription factor TCF3, which is the basis for the standard gene symbol TFPT [117]. Subsequently, TFPT was found to interact with the protein Arc and drive its nuclear import [118]. TFPT has

also been shown to stimulate apoptosis and associate with other apoptosis-inducing proteins [119, 120]. There is some evidence to suggest TFPT may function as a transcription factor. TFPT was reported to contain a b-ZIP domain [121], and BLAST analysis of the TFPT amino acid sequence reveals that it has 30% sequence identity with the transcription factor DNA damage-inducible transcript 3 protein, with particularly high homology to the leucine zipper motif in this protein. TFPT was found to possess sequence-specific DNA binding activity in a screen for DNA binding proteins using a transcription factor protein microarray [122]. Many transcription factors have been reported to be SUMO-modified, and SUMOylation in many cases has been shown to regulate their transcription factor activity. There is both indirect and direct evidence to support the SUMOylation of TFPT. TFPT was found to associate with Arc, a protein that accumulates in PML bodies, which are highly enriched with SUMOylated proteins [123, 124]. TFPT was identified as a target of SUMO2 in an affinity-purification and mass-spectrometry study [125]. Recently, TFPT was shown to be SUMO-modified at K216, which is contained within a SUMOylation consensus site [126].

3.3. The INO80 Complex Subunit INO80E

Much less is known about the INO80 complex subunit INO80E. It was found to be associated with the transcription factor FOXP3 in T cell hybridoma cells in an affinity-purification mass spectrometry screen [127]. Protein expression was found in the liver, retina and multiple immune system cell types based on a recent effort to “map” the human proteome by comprehensively identifying proteins in a variety of tissues and cell types [128].

3.4. SUMOylation in Chromatin Remodeling

A wide range of chromatin modifying enzymes and chromatin associated proteins have been found to possess SUMO-binding activity or be regulated in some way by SUMOylation [97]. The interaction between the histone deacetylase HDAC2 and the transcription factor Elk-1 was shown to be enhanced by SUMOylation, resulting in recruitment of the enzyme to DNA and subsequent histone deacetylation [99]. The SETDB1 methyltransferase was shown to interact with the SUMOylated form of the KAP-1 corepressor, and reduced KAP-1 SUMOylation was correlated with reduced SETDB1 genome occupancy and H3K9 trimethylation levels [129]. The nucleosome remodeling enzyme Mi-2 was found to bind directly to SUMO and SUMO modified Sp3, and Mi-2 target gene occupancy was dependent on Sp3 SUMOylation state [130].

Chapter 4

Development of a Protein Microarray Assay to Identify SUMO E3 Ligase-Specific Substrates

This section has been adapted from: Cox, E., Uzoma, I., Guzzo, C., Jeong, J.S., Matunis, M., Blackshaw, S., Zhu, H. (2014) Identification of SUMO E3 Ligase Specific Substrates Using the HuProt Human Proteome Microarray. *Methods in Molecular Biology* (submitted)

4.1. Abstract

The functional protein microarray is a powerful and versatile systems biology and proteomics tool that allows the rapid activity profiling of thousands of proteins in parallel. We have recently developed a human proteome array, the HuProt array, which includes ~80% of all the full-length proteins of the human proteome. In one recent application of the HuProt array, we identified numerous SUMO E3 ligase-dependent SUMOylation substrates. For many SUMO E3 ligases, only a small number of substrates have been identified and the target specificities of these ligases therefore remain poorly defined. In this protocol, we outline a method we developed using the HuProt array to screen the human proteome to identify novel SUMO E3 ligase substrates recognized by specific E3 ligases.

4.2. Introduction

The functional protein microarray is a powerful and versatile systems biology and proteomics tool that allows the rapid activity profiling of thousands of proteins in parallel. Applications of functional protein microarrays range from the identification of protein-binding properties, to surveying targets of posttranslational modifications, to uncovering novel enzymatic activities. Since the development of the yeast proteome microarray over 10 years ago [1], more recent work has seen the development of complete and near-complete proteome arrays representing viruses, bacteria and plants [2-4]. However, most existing human protein microarrays are comprised of only a minority of the human proteome [5-9]. We have recently developed a human proteome microarray, the HuProt array, which includes

nearly 20,000 full-length human proteins [10]. The proteins used to generate this microarray were expressed in yeast and purified under native conditions. Expressing recombinant eukaryotic proteins in yeast improves the likelihood that proteins will retain their biological activity relative to prokaryotic and *in vitro* expression systems.

Numerous collaborations between our labs and others have so far harnessed the power of the HuProt array to profile a wide range of protein activities. The role of posttranslational modifications in regulating enzymatic activity is one area of investigation particularly well suited for the HuProt array platform. A screen to define the S-nitrosylated proteome revealed an important regulatory role for this posttranslational modification in the control of ubiquitin E3 ligase activity [131]. In other work, phosphorylation and glycosylation states of the protein kinase CK2 were shown to affect its substrate specificity [40]. The HuProt array has also been used in two separate studies to link novel protein-RNA interactions to neurological disease, including an interaction between RNA splicing factors and a long noncoding RNA linked to schizophrenia [132] and an interaction between multiple RNA binding proteins and an expanded repeat-containing transcript implicated in amyotrophic lateral sclerosis [133]. Another ongoing project in our labs is the generation of monospecific monoclonal antibodies whose specificity can be quickly evaluated using the HuProt array [10]. The utility of the HuProt array further extends to exciting clinical applications including the identification of novel biomarkers that may be used as a diagnostic tool in primary biliary cirrhosis, an autoimmune disease of the liver [53].

Protein SUMOylation is an essential posttranslational modification in most organisms, including yeast, *C. elegans*, Arabidopsis, and mice [73]. The reversible SUMO-modification

of target proteins involves an enzymatic cascade chemically similar to ubiquitylation, involving E1 activating, E2 conjugating, E3 ligating enzymes and SUMO proteases. As the only classes of SUMOylation enzymes for which multiple members have been identified, the SUMO E3 ligases and the SUMO proteases have been proposed to be the major factors determining substrate specificity. Recently, we have conducted SUMOylation assays using the HuProt microarray to identify numerous previously uncharacterized SUMO E3 ligase-dependent substrates using a subset of human SUMO E3 ligases. While our study focused on some of the best characterized SUMO E3 ligases, recently additional SUMO E3 ligases have been described [134-138] and it is likely that new SUMO E3 ligases await discovery [139]. The methods that we describe here could be used to identify substrates for these additional SUMO E3 ligases. For most SUMO E3 ligases, only a limited number of substrates are known. In this chapter, we will describe the on-chip SUMOylation protocol that we have developed so that the reader may conduct SUMOylation assays using the HuProt microarray with their SUMO E3 ligase of interest.

4.3. Materials

4.3.1. Equipment

1. HuProt human proteome microarray (CDI Laboratories, USA).
2. Bench-top centrifuge (Thermo Scientific: HERAEUS Multifuge 3SR+ centrifuge).
3. Four-well dish (NUNC 267061).

4. Humidity chamber (USA Scientific pipet tip box with wet paper towels inside).
5. Laboratory tissues (Kimwipes).
6. LifterSlip Coverslips (Fisher: 22035809)
7. Micro slide boxes (VWR 48444-004)
8. Microarray analysis software, GenePix Pro 6.0 (MDS Analytical Technologies)
9. Orbital shaker.
10. GenePix 4000B Microarray Scanner (Molecular Devices, USA)

4.3.2. Purification of the SUMO E1 enzyme

1. E1 Binding Buffer**:** 20 mM Tris-HCl pH 8.0, 350 mM NaCl, 1mM beta-mercaptoethanol, 10 mM imidazole.
2. E1 Wash Buffer: 20 mM Tris-HCl pH 8.0, 350 mM NaCl, 1mM beta-mercaptoethanol, 20 mM imidazole.
3. E1 Elution Buffer: 20 mM Tris-HCl pH 8.0, 350 mM NaCl, 1mM beta-mercaptoethanol, 400 mM imidazole.

4.3.3. Purification of the SUMO E2 enzyme

1. E2 wash buffer: 1X PBS, 1mM DTT.

4.3.4 Common reagents for expression and purification of SUMO protein and SUMOylation enzymes

1. General lysis buffer: 1X PBS, 1mg/ml lysozyme, 2mM DTT, 1X Roche protease inhibitor cocktail (EDTA-free), 10 U/ml benzonase.
2. Enzyme dialysis buffer: 20mM Tris-HCl pH 8.0, 75 mM NaCl, 1mM □-mercaptoethanol.
3. PreScission protease (GE Healthcare)
4. IPTG
5. 1X PBS
6. lysozyme powder
7. 1M DTT
8. Roche protease inhibitor cocktail (EDTA-free)
9. Benzonase
10. Glutathione sepharose (GE Healthcare)
11. Ni-NTA agarose (Life Technologies)

4.3.5 SUMO Antibody Labeling

1. SUMO-1 affinity-purified mouse monoclonal antibody (21C7) (#33-2400, Life Technologies, USA)
2. DyLight 549 Antibody Labeling Kit (#53034, Pierce Biotechnology)

4.3.6. On-chip SUMOylation assay

1. SUMO blocking buffer: 1X TBS pH 7.4, 2% BSA, 0.05% Tween-20.

2. 2X SUMO conjugation buffer: 40 mM HEPES pH 7.3, 200 mM NaCl, 20 mM $MgCl_2$, 0.2 mM DTT
3. SUMO reaction mix (for 200ul reaction): 100ul 2X SUMO conjugation buffer, 0.25-2.5 □M E1 enzyme, 0.07-7 □M E2 enzyme, 20 nM E3 ligase (See Note 2)

4.4. Methods

4.4.1. Purification of SUMO Protein

1. Streak out colonies onto a LB + ampicillin agar plate from a glycerol stock of GST-SUMO in pGEX6p.1 in BL21 cells.
2. Pick a single colony and inoculate into 5 ml LB with ampicillin overnight at 37°C.
3. Dilute the 5 ml culture into 50 ml LB with ampicillin culture overnight at 37°C.
4. Dilute 50 ml culture into 1 L LB with ampicillin until the OD = 0.6, then drop the temperature to 20°C and induce with 1mM IPTG and shake overnight at 20°C
5. Freeze pellet at -80°C until ready to proceed with purification.
6. Thaw pellet at 37°C, resuspend with 25ml general lysis buffer, rotate at RT for 15min.
7. Centrifuge at 4°C for 30min at 14,000rpm to pellet insoluble material.

8. During centrifugation prepare glutathione sepharose by washing 2ml 50% glutathione sepharose with 1X PBS in 50ml tube. Repeat 2X.
9. Bind protein by mixing supernatant with glutathione sepharose, rotate at 4°C for 1h or longer.
10. Wash sepharose by spinning down sepharose, discard supernatant, resuspend sepharose in 1ml of 1X PBS. Transfer sepharose to column; wash with 12ml of 1X PBS.
11. Cleave SUMO from GST tag by transferring sepharose bound with GST-precission protease and GST-SUMO to the same column. Parafilm column to prevent leakage. Incubate with shaking at 4°C overnight.
12. Allow sepharose to settle. Remove supernatant containing purified untagged SUMO. Apply 4ml of 1X PBS and repeat.
13. Dialyze overnight against 2 L 1X PBS
14. Concentrate protein using a micron centrifugal filter (10kDa MWCO) by spinning 20 min @ 4000 rpm. Aliquot and freeze with liquid nitrogen and store at -80°C.

4.4.2. Purification of SUMO E1 Enzyme

1. Streak out colonies onto a an LB + ampicillin agar plate from a glycerol stock of His-hE1 (His-Aos1/Uba2) in BL21 cells
2. Pick a single colony and inoculate into 5 ml LB with ampicillin overnight at 37°C.

3. Dilute the 5 ml culture into 50 ml LB with ampicillin culture overnight at 37°C.
4. Dilute 50 ml culture into 1 L LB with ampicillin until the OD = 0.6, then drop the temperature to 20°C and induce with 1mM IPTG and shake overnight at 20°C
5. Freeze pellet at -80°C until ready to proceed with purification.
6. Thaw pellet at 37°C, resuspend with 25ml general lysis buffer, rotate at RT for 15min.
7. Centrifuge at 4°C for 30min at 14,000rpm to pellet insoluble material.
8. During centrifugation prepare Ni-NTA agarose by washing 2ml 50% Ni-NTA agarose with E1 binding buffer in 50ml tube. Repeat 2X.
9. Bind protein by mixing supernatant with Ni-NTA agarose, rotate at 4°C for 1h or longer.
10. Wash agarose by spinning down agarose, discard supernatant, resuspend agarose in 1ml of wash buffer. Transfer agarose to column; wash with 12ml of E1 wash buffer.
11. Elute with 3 ml of E1 elution buffer. Collect fractions and analyze by SDS-PAGE. Pool fractions containing protein.
12. Dialyze overnight against 2 L enzyme dialysis buffer.
13. Concentrate protein using a micron centrifugal filter (10kDa MWCO) by spinning 20 min @ 4000 rpm. Aliquot and freeze with liquid nitrogen and store at -80°C.

4.4.3. Purification of SUMO E2 Enzyme

1. Streak out colonies onto a an LB + ampicillin agar plate from a glycerol stock of GST-Ubc9 in pGEX6p.1 in BL21 cells
2. Pick a single colony and inoculate into 5 ml LB with ampicillin overnight at 37°C.
3. Dilute the 5 ml culture into 50 ml LB with ampicillin culture overnight at 37°C.
4. Dilute 50 ml culture into 1 L LB with ampicillin until the OD = 0.6, then drop the temperature to 20°C and induce with 1mM IPTG and shake overnight at 20°C
5. Freeze pellet at -80°C until ready to proceed with purification.
6. Thaw pellet at 37°C, resuspend with 25ml general lysis buffer, rotate at RT for 15min.
7. Centrifuge at 4°C for 30min at 14,000rpm to pellet insoluble material.
8. During centrifugation prepare glutathione sepharose by washing 2ml 50% glutathione sepharose with 1X PBS in 50ml tube. Repeat 2X.
9. Bind protein by mixing supernatant with glutathione sepharose, rotate at 4°C for 1h or longer.
10. Wash sepharose by spinning down sepharose, discard supernatant, resuspend agarose in 1ml of E2 wash buffer. Transfer sepharose to column; wash with 12ml of E2 wash buffer.

11. Cleave UBC9 from GST tag by transferring beads bound with GST-precission protease and GST-UBC9 to the same column. Parafilm column to prevent leakage. Incubate with shaking at 4°C overnight.
12. Allow sepharose to settle. Remove supernatant containing purified untagged UBC9. Apply 4ml of E2 wash buffer and repeat.
13. Dialyze overnight against 2 L enzyme dialysis buffer.
14. Concentrate protein using a micron centrifugal filter (10kDa MWCO) by spinning 20 min @ 4000 rpm. Aliquot and freeze with liquid nitrogen and store at -80°C.

4.4.4. SUMO Antibody Labeling

1. Add 40ul borate buffer to 0.5 ml of 0.5 mg/ml affinity-purified SUMO-1 antibody.
2. Add 0.5 ml of antibody in borate buffer to the vial of DyLight reagent and vortex gently.
3. Briefly centrifuge to collect the sample in the bottom of the tube.
4. Incubate the reaction mixture for 60 minutes at room temperature protected from light.
5. Mix purification resin to ensure uniform suspension and add 400 ul of the suspension into both spin columns. Centrifuge for 45 seconds at 1000 g to remove the storage solution. Discard the used collection tubes and place the columns in new collection tubes.

6. Add 250-270 □l of the labeling reaction to each spin column and mix the sample with the resin by vortexing.
7. Centrifuge columns for 45 seconds at 1000 g to collect the purified proteins. Combine the samples from both columns (0.5 ml total).
8. Aliquot and store the labeled antibody at -20°C.

4.4.5. On-chip SUMOylation assay

1. Rinse arrays by quickly dunking in a beaker of 300 mL of TBS.
2. Place each array in a well of a four-well dish with 3 mL of SUMO blocking buffer per well.
3. Block protein microarray by gently shaking overnight at 4°C.
4. Prepare the reaction mix and keep on ice. Add E1 and E2 enzymes and E3 ligase (optional) immediately before the end of the blocking step.
5. Remove arrays from blocking buffer and carefully wick off liquid by tapping the edge on a paper towel and place arrays in a humidity chamber (See Note 1).
6. Add SUMO reaction mix to each slide carefully and place coverslip on top, being careful to avoid bubbles.
7. Incubate at 37°C for 90 minutes (depending on enzyme activity).
8. Immediately after start of incubation, pre-warm appropriate volume of 1% SDS to 55°C for later washing steps.
9. Remove coverslip by gently sliding off array.

10. Place arrays in a large empty pipette tip box (e.g. a 1000ul tip pipette tip box works well) and wash gently on orbital shaker 3x for 10 minutes at room temperature with enough TBST to completely cover the arrays.
11. Wash 3 times with 1% SDS warmed to 55°C for 5 minutes.
12. Wash once with ddH_2O.
13. Dilute labeled SUMO-1 antibody in blocking buffer at 1:1000 dilution. Apply 200 µL of the antibody mixture to each array and cover with LifterSlip. Place arrays in humidity chamber and incubate slides with labeled SUMO-1 antibody for 1 hour at room temperature.
14. Remove coverslip by gently sliding off array and wash slides 3 times for 10 minutes in 50-100 ml TBST in a clean pipette tip box.
15. Wash the slides once in 50-100 ml milliQ water for 5 minutes to remove residual salts from the surface of the microarray.
16. Place each array horizontally into a micro slide box with laboratory tissues on the bottom. Centrifuge the box in a benchtop centrifuge for 3 min at 2000 rpm.
17. Scan the microarray with a GenePix 4000B scanner with 5-µm resolution detection at 532 nm. Ensure that no spots appear as saturated signals (saturated signal appears as white). Save the scanned images as TIF files. All slides that will be compared should be scanned using the same gain and power settings.

4.5. Notes

1. The humidity chamber is made by placing a folded paper towel in the bottom of an empty pipette tip box, adding one inch of ddH_2O, and replacing the tip holder on the box. The arrays will sit on the tip holder and be covered with the lid.
2. SUMOylation of many proteins occurs in the absence of E3 ligases at relatively high E1 and E2 concentrations, therefore: E1 and E2 concentrations should be used in the low end of this range when using an optional E3 ligase. When it is desired to omit an E3 ligase, it is recommended to use E1 and E2 concentrations at the high end of this range.

Chapter 5

Global Analysis of SUMO-Binding Proteins Identifies SUMOylation as a Key Regulator of the INO80 Chromatin Remodeling Complex

5.1. Abstract

SUMOylation is an essential posttranslational modification in most organisms that is thought to function through its ability to modulate the protein-protein interactions of a SUMO target protein. Accordingly, the function of SUMOylation can be better understood through the identification of SUMO-modified targets as well as downstream SUMO-interacting proteins. In order to identify novel SUMO-interacting proteins, we developed a SUMO-binding assay using the human proteome microarray, which includes nearly 20,000 full-length human proteins. We then integrated SUMO-binding with a SUMOylation data set recently generated in our laboratory, as well as protein-protein interaction data from publicly available databases to perform network motif analysis. We focused on a single network motif we termed a SUMOmod PPI (SUMO-modulated Protein-Protein Interaction) that included the INO80 chromatin remodeling complex subunits TFPT and INO80E. We validated the SUMO-binding activity of INO80E and that TFPT is a SUMO substrate both *in vitro* and *in vivo.* We then went on to demonstrate a key role for SUMOylation in mediating the interaction between these two proteins, both *in vitro* and *in vivo*. By demonstrating a key role for SUMOylation in regulating the INO80 chromatin remodeling complex, this work illustrates the power of bioinformatics analysis of large datasets in predicting novel biological phenomena.

5.2. Introduction

The functional protein microarray is a powerful and versatile systems biology and proteomics tool that allows the profiling of the activity of thousands of proteins in parallel. Applications of functional protein microarrays range from the profiling of protein interaction, discovery of posttranslational modifications, and identification of novel enzymatic activities. Since the development of the yeast proteome microarray over 10 years ago [1], more recent work has seen the development of complete and near-complete proteome arrays representing viruses, bacteria and plants [2-4]. However, most existing human protein microarrays are comprised of only a minority of the human proteome [5-9]. We have recently developed a human proteome microarray, the HuProt array, which includes nearly 20,000 full-length human proteins [10]. The proteins used to generate this microarray were expressed in yeast and purified under native conditions. Expressing recombinant eukaryotic proteins in yeast improves the likelihood that proteins will retain their biological activity relative to prokaryotic and *in vitro* expression systems.

Numerous studies have harnessed the power of the HuProt array to profile a wide range of protein activities, including RNA-protein interactions[132, 133], analysis of monoclonal antibody specificity [10] and serum profiling [53], and identification of substrates of protein kinases [40] and S-nitrosylation [131].

Recently, we have conducted SUMOylation assays using the HuProt microarray to identify numerous previously uncharacterized SUMO E3 ligase-dependent substrates using a

subset of human SUMO E3 ligases. Protein SUMOylation is an essential posttranslational modification in most organisms, including yeast, *C. elegans*, Arabidopsis, and mice [73]. The reversible SUMO-modification of target proteins involves an enzymatic cascade chemically similar to ubiquitylation, involving E1 activating, E2 conjugating, E3 ligating enzymes and SUMO proteases. As the only classes of SUMOylation enzymes for which multiple members have been identified, the SUMO E3 ligases and the SUMO proteases have been proposed to be the major factors determining substrate specificity.

The ultimate impact of SUMOylation is thought to arise from its ability to modulate the protein-protein interactions of a SUMO target. Accordingly, the function of SUMOylation is determined through a combination of the identification of SUMO-modified targets as well as downstream SUMO-interacting proteins. In order to identify novel SUMO-interacting proteins, we developed a SUMO-binding assay using the human proteome microarray. Previous efforts to identify SUMO-binding proteins have found relatively small numbers of proteins [96, 104, 112]. In addition, these studies were limited in that they either looked at proteins expressed in a limited number of tissue types, or that they were restricted to the identification of proteins localized to the nucleus that bound to only a single SUMO isoform. The studies also relied on the use of MS-pulldowns, which are often unable to reliably detect low-abundance proteins. For our studies, we utilized the HuProt microarray containing nearly 20,000 purified human proteins, thus avoiding bias in favor of particular tissue type or subcellular compartment, and allowing identification of low-abundance target proteins. We also looked at binding to both SUMO1 and SUMO2 monomers, as well as SUMO1 and SUMO2 trimers, which were used as model SUMO chains. Our dataset

represents an over 10-fold increase in the number of known SUMO-interacting proteins. The vast majority of these proteins had not been previously reported as having SUMO binding activity.

To use our data set to make novel predictions about SUMOylation function, we integrated our SUMO-binding and SUMOylation data with protein-protein interaction data from publicly available databases to perform network motif analysis. We identified 21 three-component network motifs including 9 network motifs that were significantly enriched and 4 network motifs that were significantly depleted. We focused on a single network motif containing a SUMO-binding protein and a SUMO-modified protein which were previously known to interact based on the literature, along with SUMO itself. This specific network motif suggests a potential role for SUMO in modulating the protein-protein interaction between the other 2 nodes, and we thus termed this motif a SUMOmod PPI (SUMO-modulated Protein-Protein Interaction). We then validated an example of this network motif that comprised INO80 chromatin remodeling complex subunits INO80E and TFPT, along with SUMO2. We found that TFPT could be SUMOylated in transfected cells and identified the relevant SUMOylation site on the protein. We also validated the SUMO-binding activity of INO80E and identified the region of INO80E important for this interaction. Then we went on to demonstrate that SUMO helps facilitate the interaction between INO80E and TFPT. This work demonstrates the power of bioinformatics analysis of large datasets in predicting novel biological phenomena.

5.3. Results

Previous work reported that the first alpha helix and second beta strand of both SUMO1 and SUMO2 can act as a SUMO paralog recognition region that interacts with the SIMs of SUMO paralog-specific binding proteins [86, 106, 140]. Although only a handful of SUMO paralog-specific binding proteins have been identified to date, structural differences in this region suggest that more paralog-specific binding proteins remain to be discovered. In order to identify novel SUMO paralog-specific binding proteins we used purified SUMO1 and SUMO2 monomers as well as SUMO1 and SUMO2 trimers as model SUMO chains (Figure 5A) that had been previously used to identify SUMO2-specific binding proteins [86]. We then used these protein probes in a novel human proteome microarray-based SUMO-binding assay (Figure 5B). For detection of SUMO binding proteins on the proteome microarray, we prepared fluorescently labeled SUMO-isoform specific monoclonal antibodies. We then performed SUMO-binding assays in triplicate for each of the 4 probes. Among 461 total SUMO binding proteins that were identified by our assay, only 8 had been previously identified in other work (Figure 11). 6 of these 8 previously identified SUMO-binding proteins were found by a SUMO2-focused affinity purification and mass spectrometry based study and include ANXA1, HRNRPK, RUVBL1, TRIM26, ZBED1 and ZNF451 [112]. The 4 SUMO protein probes that we tested showed wide variation in their binding specificity (Figure 6A). The SUMO2 monomer probe bound to the largest number of proteins (306) as well as the largest number of unique proteins (128). The SUMO1 trimer bound 197, the SUMO1 monomer bound 183, and the SUMO2 trimer bound 139 proteins in total. The pairwise probe combinations of SUMO1 trimer and SUMO2 monomer, and

SUMO1 monomer and SUMO2 monomer, each bound the largest and same number of proteins (108). On the other hand the SUMO1 trimer and SUMO2 trimer bound to the smallest number of shared proteins (66). 39 proteins showed promiscuous SUMO-binding activity and interacted with all 4 SUMO probes.

We next conducted gene ontology analysis on the novel SUMO binding proteins (Figure 6B). This analysis revealed several notable properties of these proteins. With regard to cellular component, we observed significant enrichment for cytoplasmic localization among proteins that bound by all four probes.

Figure 5. (A) SUMO Probes and SUMO-binding assay. (Top) SUMO1 and SUMO2 protein surfaces generated using MacPyMOL with previously reported crystal structures of SUMO1 and SUMO2, PDB ID 2PE6 and 4BKG, respectively[141, 142]. The surfaces corresponding to the first alpha-helix and second beta-strand for both SUMO1 and SUMO2 are shown in color (nitrogens are shown in red, oxygens in blue and carbons in green). This surface has been reported to be important for interactions with other proteins.
(Bottom) Cartoon showing 4 probes that were used in our experiments. The SUMO1 and SUMO2 trimer probes, described previously[86], were expressed as trimeric fusion proteins. **(B)** Schematic of the SUMO-binding protocol used with the human proteome microarray. **(C)** Representative chip images showing human proteome microarray visualized with an antibody to GST, showing all proteins (left, in green), and an antibody to SUMO2 following a SUMO-binding experiment with the SUMO2 trimer (right, in red)

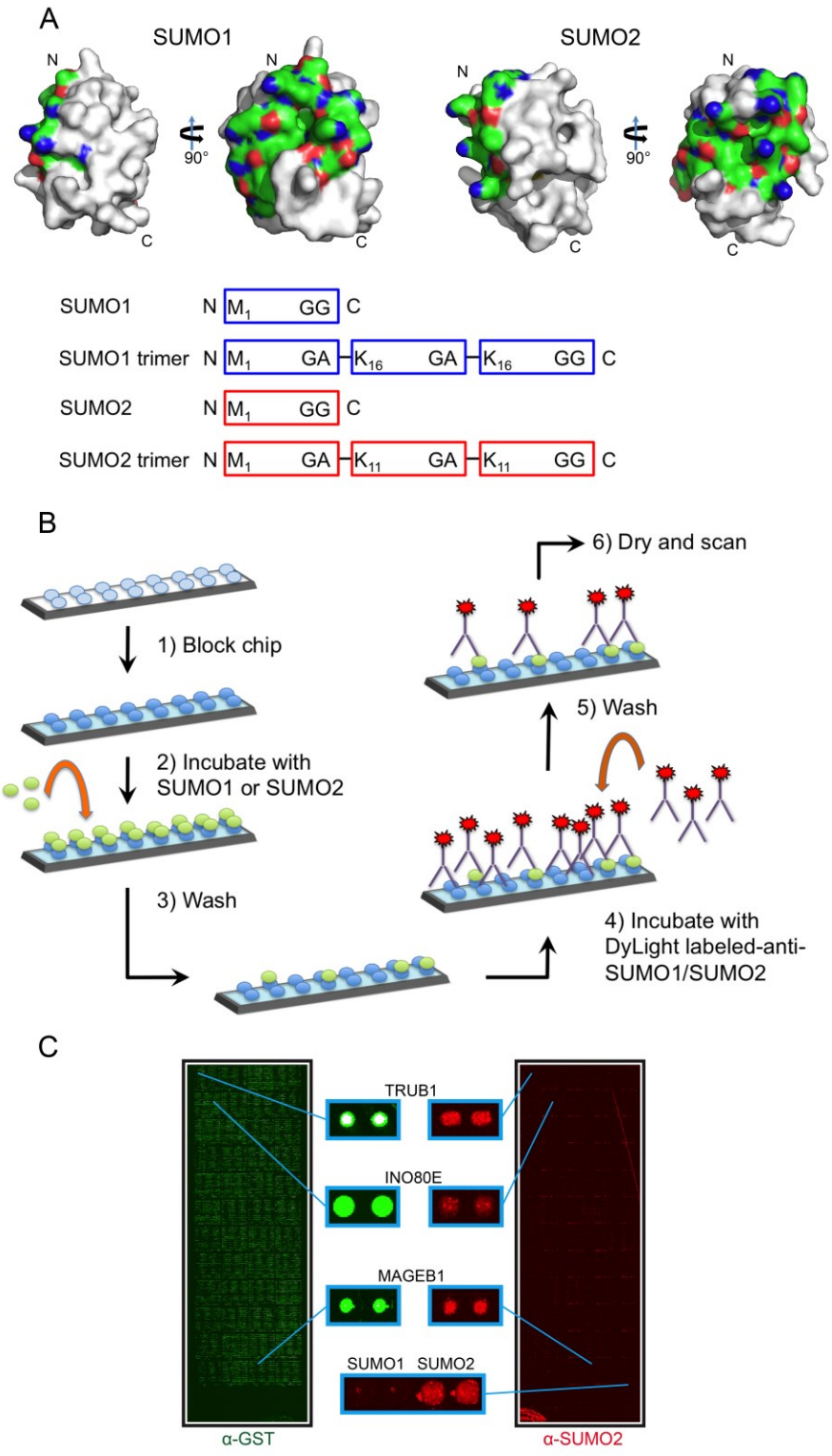

Figure 5. SUMO Probes and SUMO-binding assay

Considering that many previously identified SUMO-modified proteins are localized to the nucleus, in particular transcription factors, this enrichment of cytoplasmic proteins suggests that the importance of SUMOylation in the cytoplasm may be underappreciated. We also found that some GO categories were enriched among proteins that bound to all 4 probes or a single probe as well as different combinations of 2 or 3 probes. For example, proteins associated with the GO category of "actin binding", "catalytic activity" and "oxidoreductase activity" were enriched among targets bound by all 4 probes. A particularly surprising result of this analysis was the finding that of all pairwise combinations of probes tested, proteins that bound the SUMO1 trimer and SUMO2 monomer, which are two of the most dissimilar probes, shared the largest number of molecular function GO categories. With regard to biological process, enriched GO categories common to proteins that bound all 4 probes included 3 terms: "metabolic process", "muscle contraction" and "oxidation reduction". As with molecular function, the SUMO1 trimer and SUMO2 monomer binding proteins shared the largest number of GO categories compared to any other pairwise probe combination.

We hypothesized that by integrating our SUMO-binding data with recently published SUMOylation data [143] as well as protein-protein interaction data from publicly available databases, we could make novel predictions about SUMOylation function. Using a network motif analysis based bioinformatics approach, we generated a list of 21 tripartite network motifs, identifying 9 enriched motifs and 4 depleted motifs (Figure 7). Each motif is associated with 2 parameters, including *n*, the number of examples of the particular motif type that we identified, and *Z*, an enrichment score. This enrichment score is calculated as the difference between the observed occurrence of this network motif in the integrated

SUMOylation network and its average occurrence in several hundred random networks, normalized with the standard deviation [144]. In these network motifs, at least one node represents SUMO, and at least one node represents a protein from either the SUMO-binding or SUMOylation datasets. These 3 proteins are then connected by edges that represent one of three types of protein-protein interaction, including 1) noncovalent SUMO-binding identified in our assay or 2) covalent SUMOylation, or 3) a protein-protein interaction from one of several publicly available databases.

We then selected one network motif that included two proteins previously reported to interact that included a novel SUMO-binding protein (our dataset) and a SUMOylated protein [143]. This network motif immediately suggests a potential role for SUMO in modulating this protein-protein interaction, thus we termed this network motif a "SUMO-modulated protein-protein interaction" (SUMOmod PPI). One such SUMOmod PPI consisted of the INO80 chromatin remodeling complex subunits INO80E and TFPT, and SUMO2. Our proteome microarray assay revealed that INO80E binds specifically to the SUMO2 monomer and the SUMO2 trimer.

Figure 6. SUMO-binding assay results and gene ontology analysis.

(A) Venn diagram showing the number of SUMO-binding experiment hits

(B) Gene ontology analysis on proteins identified in our screen for SUMO-binding proteins. Gene ontology categories that are significantly enriched among proteins that bound to one of our 4 probes are shown as a colored box, color-coded by probe.

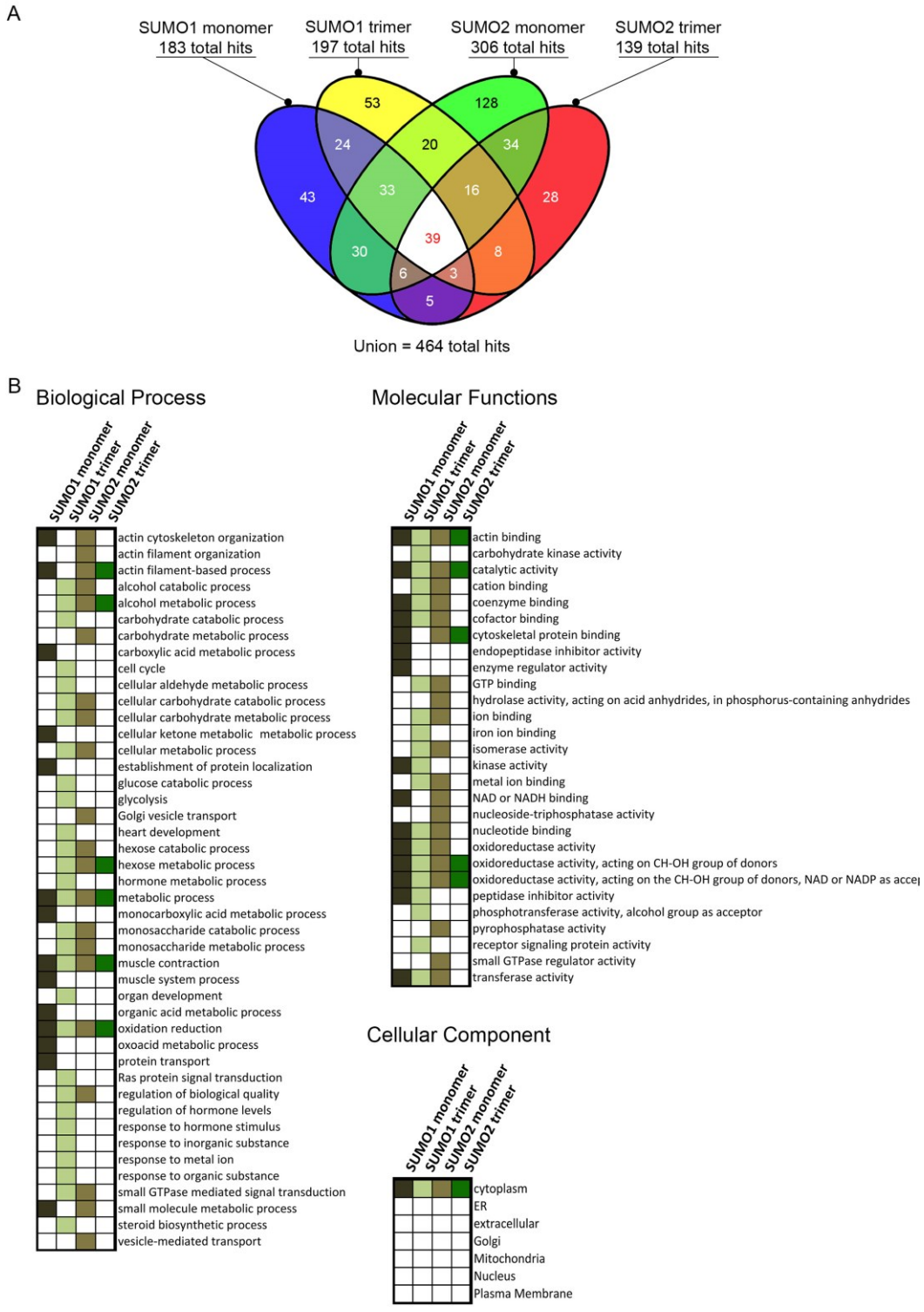

Figure 6. SUMO-binding assay results and gene ontology analysis.

On the other hand, the other protein, TFPT, was previously found to be specifically modified by SUMO2 in the presence of the SUMO E3 ligases PIAS1 and PIAS3 [143]. As a candidate "SUMOmod PPI" we hypothesized that SUMO may play a role in modulating the interaction between INO80E and TFPT. Before we could proceed with testing our hypothesis that SUMOylation may modulate the interaction between the proteins INO80E and TFPT, we first wanted to validate their SUMO-binding and SUMOylation properties, respectively, using conventional assays.

First, we evaluated the binding of INO80E translated *in vitro* in rabbit reticulocyte lysate in the presence of [^{35}S] methionine, to immobilized GST-tagged SUMO-1 or SUMO2 polymers. We observed a strong interaction between INO80E and the SUMO2 polymer in contrast to a very weak, almost undetectable interaction between INO80E and the SUMO-1 polymer (Figure 12), in agreement with our human proteome microarray data.

Visual inspection of the amino acid sequence of INO80E revealed a canonical SUMO-interaction motif (SIM) at the protein C-terminus containing a hydrophobic core "VIDI" preceded by 3 negatively charged aspartate residues (Figure 8A). In order to test whether this predicted SIM was important for SUMO binding, we made 2 mutant forms of the INO80E protein. In one mutant, we mutated the hydrophobic residues in the hydrophobic core of the SIM to alanines (mtSIM).

Figure 7) Network motif analysis. Schematic of network motif analysis approach showing the 3 data sets that were used and all enriched and depleted network motifs. The "SUMOmod PPI" network motif that was the focus of subsequent experiments is boxed in orange.

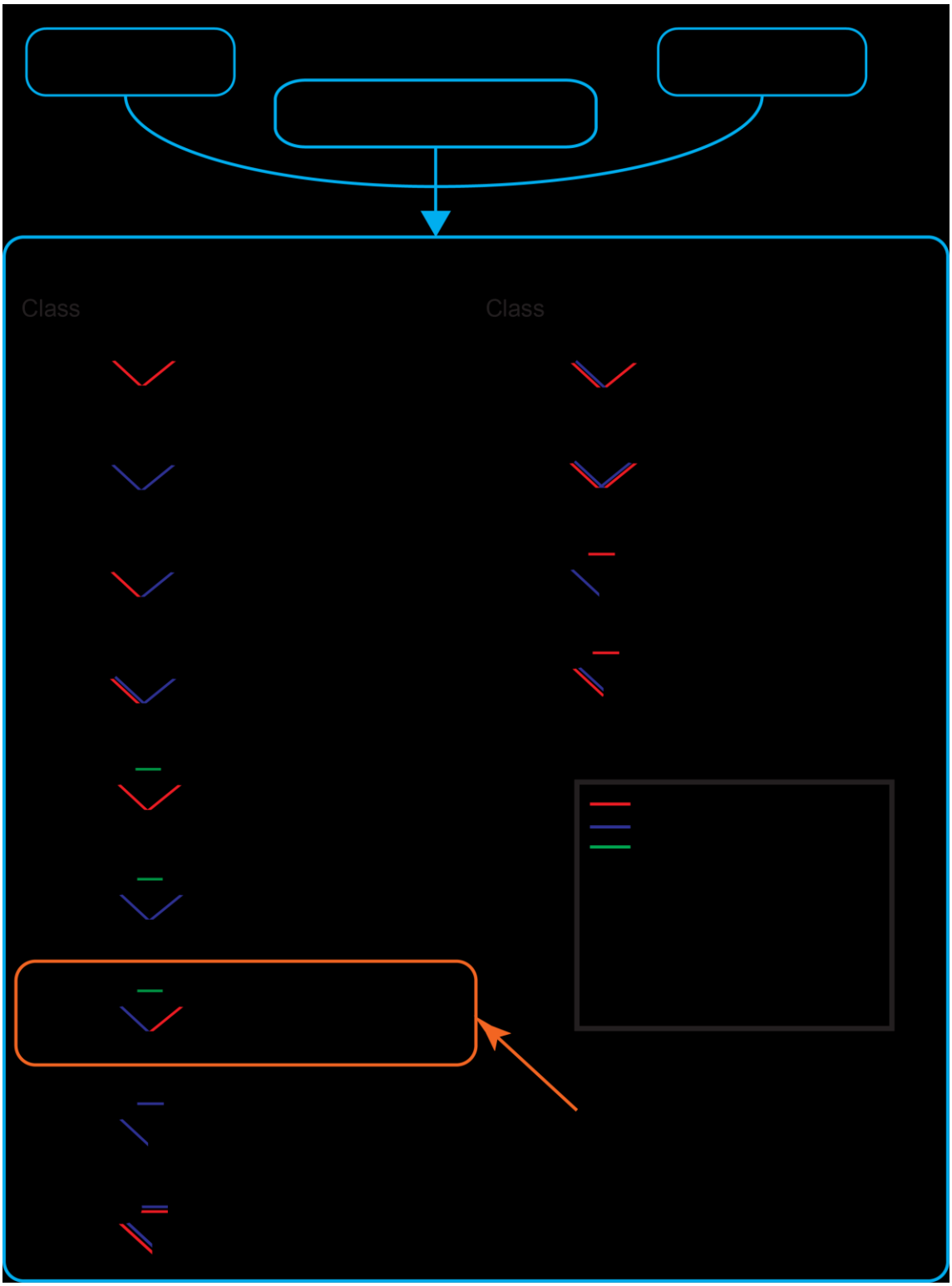

Figure 7. Network motif analysis.

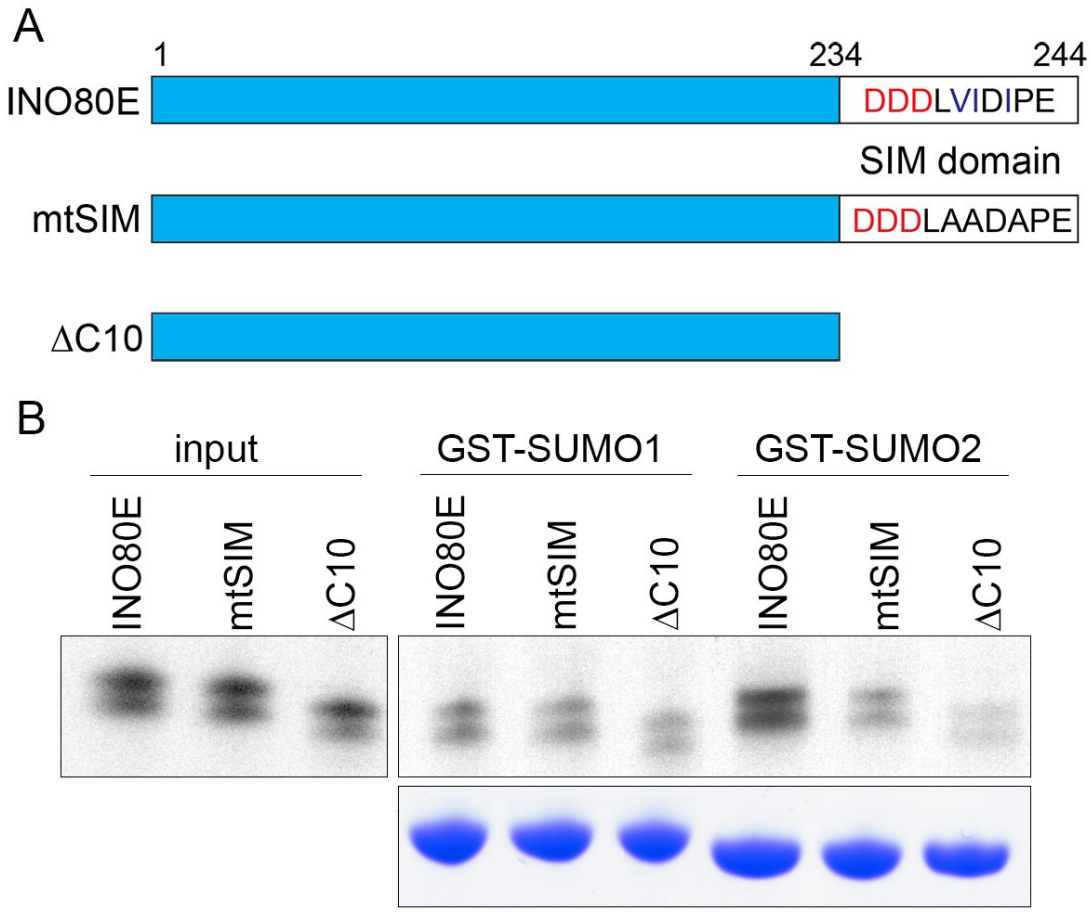

Figure 8. INO80E interacts with SUMO *in vitro*

(A) Cartoon showing INO80E, the location of the SUMO interaction motif (SIM) at the C-terminus of the protein and the 2 INO80E. The SIM is color-coded with acidic residues shown in red and hydrophobic residues in blue. Note "INO80E SIM" mutant with hydrophobic residues mutated to alanines. The "INO80EΔC10" mutant is shown with the 10 C-terminal amino acids deleted, including acidic and hydrophobic amino acids of the SIM. **(B)** Results from *in vitro* binding assay. INO80E, the INO80E SIM mutant and the INO80EΔC10 were *in vitro* translated in the presence of [^{35}S]-methionine and incubated with glutathione sepharose-bound GST-SUMO1 and GST-SUMO2. Following binding, samples

were subject to SDS-PAGE and gels were first Coomassie stained to show equal amounts of GST-SUMO1 and GST-SUMO2, then dried and exposed to film to detect radiolabeled protein.

In the second mutant, we made a truncation mutant that lacked the 10 C-terminal residues including the hydrophobic core sequence as well as the preceding negatively charged residues (ΔC10). The mtSIM mutant showed a strong reduction in binding of INO80E to GST-SUMO2, with only a slight reduction in binding to GST-SUMO1 (Figure 8B). However, while the INO80E truncation mutant (ΔC10) lacking the 10 C-terminal amino acids showed a modest reduction in binding to GST-SUMO1, it showed a dramatic reduction in binding to GST-SUMO2, approaching the lower limit of detection in our assay. This result demonstrates the importance of both the charged and hydrophobic residues in the interaction between INO80E and SUMO, both for SUMO1 and, particularly, for SUMO2.

Next, we turned to the other protein in this network, TFPT. Prior work in our laboratory had identified TFPT as being modified by SUMO2 in the presence of the SUMO E3 ligases PIAS1 and PIAS3 using a human proteome microarray based assay [143]. Additionally, TFPT had been previously shown by others to be modified by SUMO2 in an MS-pulldown experiment in HeLa cells [145]. We wanted to validate that TFPT could be SUMO-modified *in vitro* and also identify the lysine residue to which SUMO was attached. Amino acid sequence analysis using GPS-SUMO software [146] revealed the presence of a consensus SUMOylation motif near the C-terminus of the protein at K216. To test whether this lysine residue was important for SUMOylation we mutated the lysine residue to arginine

(K216R). Then we conducted in vitro SUMOylation assays by expressing TFPT in rabbit reticulocyte lysate in the presence of [^{35}S] methionine followed by incubation with or without the E1 and E2 SUMOylation enzymes with either SUMO-1 or SUMO-2 (Figure 9A). When wildtype TFPT was incubated with the SUMOylation enzymes and either SUMO1 or SUMO2, we observed a strong band representing a 20kD molecular weight shift consistent with a SUMO-modified form of TFPT. However, when we incubated the in vitro translated TFPT K216R mutant with the SUMOylation enzymes and either SUMO1 or SUMO2, no such shifted band was observed. This suggests that K216 is the relevant lysine residue for TFPT SUMO-modification by both SUMO isoforms SUMO1 and SUMO2.

Next, we tested the ability of TFPT to be SUMOylated in transfected mammalian cells. We co-transfected V5-tagged TFPT with or without myc-tagged SUMO2 into HEK 293T cells, then immunoprecipitated V5-TFPT using an anti-V5 antibody, followed by immunoblotting using an anti-myc antibody to detect SUMOylated TFPT. In the absence of myc-SUMO2, no bands were observed in the anti-myc western blot (Figure 9B). However, when we co-transfected SUMO2, we saw a strong band representing a SUMO-modified TFPT. Next, we tested the ability of the TFPT K216R mutant to be SUMOylated in mammalian cells. Although TFPT K216R was expressed (see input V5 blot) at equal levels and immunoprecipitated (see IP:V5/IB:V5 blot) with equal efficiency, we failed to see any signal in the anti-myc blot, thus indicating that TFPT K216R was not SUMOylated. This shows that K216 is essential for SUMOylation of TFPT in mammalian cells.

Previously, our group had shown significant enhancement of TFPT SUMOylation in the presence of the SUMO E3 ligases PIAS1 and PIAS3 using human proteome microarray based SUMOylation assays [143]. In order to test whether PIAS1 and PIAS3 enhanced SUMOylation of TFPT in mammalian cells, we co-transfected either FLAG-tagged PIAS1 or untagged PIAS3. When we co-transfected TFPT, SUMO2 and PIAS1, we observed a strong signal corresponding to mono-SUMOylated TFPT with several additional bands representing several higher molecular weight forms of SUMOylated TFPT . When transfecting the TFPT K216R mutant with SUMO2 and PIAS1, the anti-myc immunoblot showed only weak signal corresponding to the highest range of molecular weights.

Finally we tested the ability of TFPT SUMOylation to be enhanced by the SUMO E3 ligase PIAS3. As with PIAS1, we saw a strong band representing mono-SUMOylated TFPT as well as a laddering band pattern representing several higher molecular weight forms of SUMOylated TFPT. This pattern is not observed when transfecting the TFPT K216R mutant. Our results indicate that TFPT is SUMOylated on K216 in mammalian cells, and that the SUMO E3 ligases PIAS1 and PIAS3 can stimulate formation of higher molecular weight forms of SUMOylated TFPT. We are unable to distinguish whether these high molecular weight bands represent the attachment of SUMO2 chains to K216 on TFPT, or alternatively, whether SUMOylation at K216 is a necessary event that precedes SUMOylation or even other posttranslational modifications at other sites on TFPT. INO80E and TFPT are subunits of the human INO80 chromatin-remodeling complex and interact with the N-terminal domain of the INO80 ATPase [115]. The SUMOmod PPI network motif containing INO80E, TFPT and SUMO2 predicts that SUMOylation may enhance interaction between INO80E

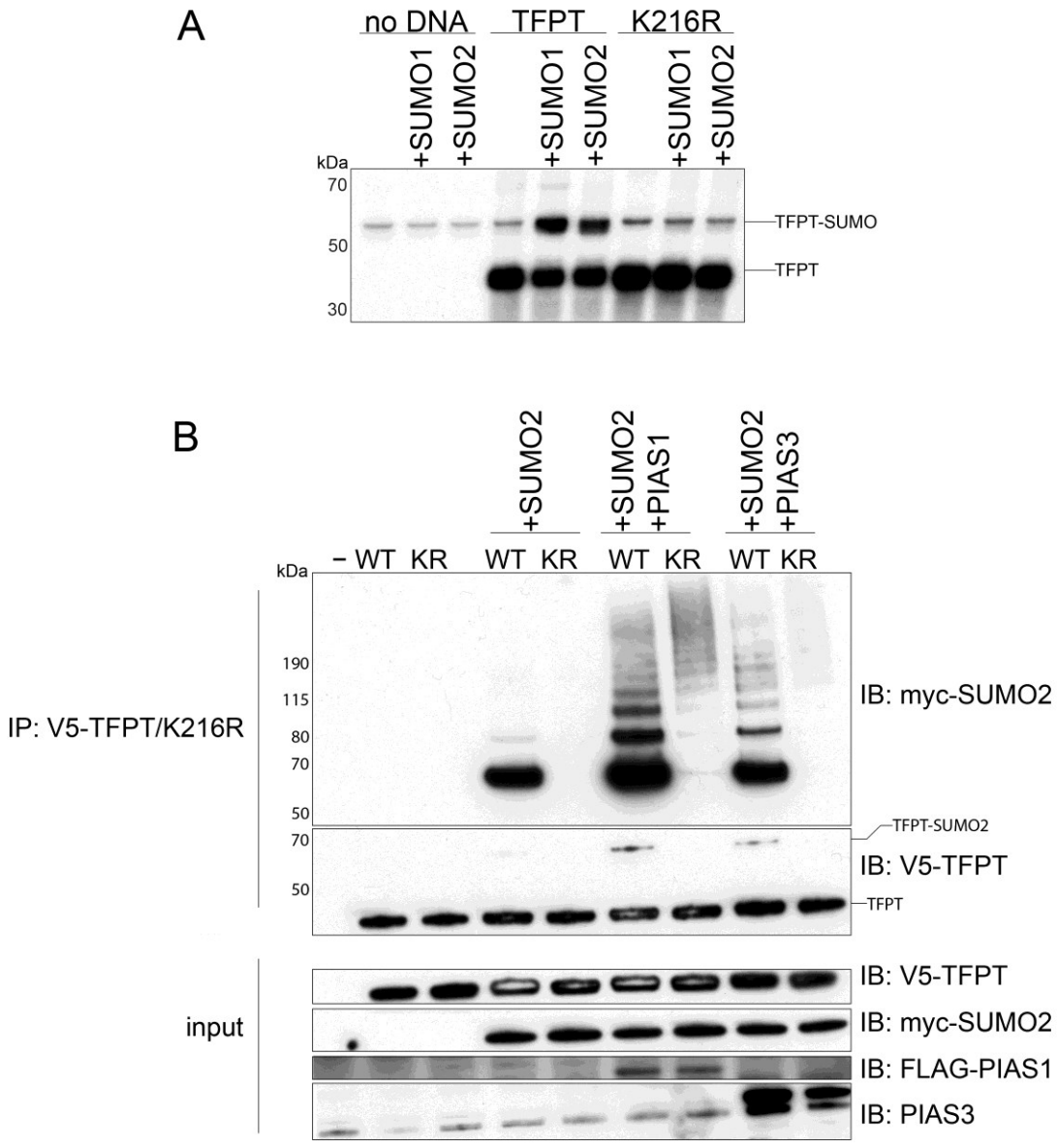

Figure 9. TFPT is SUMOylated *in vitro* and *in vivo.*

(A) Results from i*n vitro* SUMOylation assay for TFPT. Wildtype TFPT and the K216R mutant were *in vitro* translated in the presence of [^{35}S]-methionine and incubated with our without the E1 and E2 SUMOylation enzymes and either SUMO1 or SUMO2.

(B) Results from the i*n vivo* SUMOylation assay for TFPT. HEK 293T cells were co-transfected with TFPT and the K216R mutants with or without SUMO2 and either the SUMO E3 ligases PIAS1 or PIAS3. We immunoprecipitated V5-tagged TFPT and immunoblotted against myc-tagged SUMO2 to detect SUMOylated TFPT.

and TFPT. Considering that the INO80E C-terminal SUMO interaction motif is important for the ability of INO80E to interact with SUMO2, and that TFPT is modified by SUMO2, we hypothesized that the INO80E SIM may mediate the interaction between INO80E and SUMOylated TFPT. In order to test this hypothesis, we conducted an *in vitro* binding assay to allow us to measure binding between INO80E and SUMO-modified TFPT. We first *in vitro* translated TFPT in the presence of [^{35}S]-methionine, then incubated the reaction either with or without SUMO1 or SUMO2 and the E1 and E2 SUMO ligases, to give us a mixture of unmodified and SUMO-modified TFPT (Figure 10A). We then used this protein mixture to conduct binding assays with purified recombinant GST-INO80E bound to glutathione sepharose. Although the mixture of TFPT and SUMO2-modified TFPT forms contained a larger fraction of the unmodified form than the SUMOylated form, following binding to GST-INO80E, we almost exclusively recovered the SUMO2-modified form of TFPT, suggesting that INO80E preferentially interacts with the SUMOylated form of TFPT.

When we conducted a binding reaction using the INO80E ΔC10 mutant lacking the C-terminal SIM, we saw a reduction in the recovery of SUMOylated TFPT. We then repeated this experiment in triplicate and found a significant quantitative reduction in binding of the ΔC10 mutant to the SUMO2-modified form of TFPT, while binding to the SUMO1-

modified form of TFPT was not significantly reduced (Figure 10B). Additionally, the ΔC10 mutant did not show any difference in its interaction with unmodified TFPT, relative to wildtype INO80E (data not shown). This suggests that the C-terminal SIM of TFPT is specifically required for binding to the SUMO2-modified form of TFPT.

Next, we wanted to determine whether the INO80E C-terminal SIM was important for the interaction between INO80E and TFPT in mammalian cells. We conducted co-immunoprecipitation experiments in HEK293T cells to look at the interaction between INO80E and TFPT. We transfected cells with myc-INO80E and V5-TFPT followed by immunoprecipitation using the anti-myc antibody and western blot using anti-V5 showed that INO80E could co-immunoprecipitate both the wildtype and K216R mutant forms of TFPT (Figure 6C). However, when we co-transfected the INO80E ΔC10 mutant with TFPT, despite equal levels of expression of the INO80E mutant relative to the wildtype form and equal immunoprecipitation efficiency, we observed a greatly reduced ability of this mutant form of INO80E to co-immunoprecipitate TFPT. We were unable to detect the SUMO-modified form of TFPT in the co-immunoprecipitated fraction, likely due to the low stoichiometry of this form in transfected cells. In mammalian cells it appears that the interaction between INO80E and TFPT may be dependent on the INO80E SIM regardless of the SUMOylation state of TFPT. This could indicate that the INO80E SIM may be required for interaction with unmodified TFPT or alternatively, with another endogenous SUMO-modified protein that may mediate the interaction between INO80E and TFPT.

Finally we co-transfected INO80E wildtype and mutant forms with TFPT and SUMO2. We observed a rescue of the interaction between INO80E and TFPT specific to the interaction between wildtype INO80E and TFPT that was not observed when co-transfecting INO80E delta C10 and TFPT, suggesting that SUMO2 stimulated the INO80E-TFPT interaction and that the INO80E SIM was important for this effect.

We also looked at the interaction between TFPT and the ATPase subunit of the INO80 complex, INO80. Previously, it was found that both INO80E and TFPT interacted with an N-terminal region of INO80 [116]. We co-transfected a FLAG-tagged N-terminal region of INO80, V5-TFPT, and myc-SUMO2 into HCT116 cells and immunoprecipitated INO80, followed by immunoblot analysis with antibodies to myc and V5 (Figure 13). We found that the fraction of TFPT that was co-immunoprecipitated by INO80 was strongly enriched for SUMOylated TFPT. This demonstrated that SUMOylated TFPT was a likely component of the INO80 chromatin remodeling complex.

5.4. Discussion

We have identified that the interaction between the INO80 chromatin remodeling complex subunits INO80E and TFPT is modulated by SUMO. This suggests a potential mode of regulating the associated INO80 chromatin remodeling complex. Although the roles of the INO80E and TFPT subunits in modulating the INO80 complex are unknown, they have been proposed to serve a regulatory function that is likely specific to metazoans, as there are no known orthologs in yeast [116]. We confirmed the results of a human proteome

microarray-based screen to identify SUMO E3 ligase-dependent SUMOylation substrates that showed that the SUMOylation of TFPT is enhanced by the SUMO E3 ligases PIAS1 and PIAS3 [143]. We also identified the SUMOylation site at K216. After we identified this site, an effort to map SUMOylation sites proteome-wide in HeLa cells confirmed K216 as the SUMOylation site on TFPT [126]. We found that INO80E binds to both SUMO1 and SUMO2, but a C-terminal SUMO interaction motif appears to be particularly important for its interaction with SUMO2. Furthermore, INO80E shows a strong preference for interacting with SUMO2-modified TFPT over unmodified TFPT *in vitro* and this interaction is quantitatively reduced following mutation of the INO80E SIM. Interestingly, PIAS1 and PIAS3 have been identified as upstream regulators of chromatin remodeling through their role in stimulating SUMOylation of the protein MBD1, thus effecting its ability to interact with the histone methyltransferase SETDB1 [147]. Thus, PIAS1 and PIAS3-mediated SUMOylation of chromatin-associated proteins may be a general mechanism by which chromatin remodeling complex assembly is regulated. Together, our work suggests a model by which PIAS1 and PIAS3 SUMO E3 ligases promote SUMOylation of TFPT, which stimulates association with INO80E in the process of INO80 complex assembly.

There is some evidence to suggest TFPT may function as a transcription factor. TFPT was reported to contain a b-ZIP domain [115], and it was found to possess sequence-specific DNA binding activity in a screen to identify DNA binding proteins using a transcription factor protein microarray [8]. The transcription factor YY1, a known SUMO target, was shown to recruit the INO80 complex to specific gene promoters [92, 148]. TFPT may play a similar role, by binding to both the genome and to the INO80 complex in order to recruit the complex to specific genomic locations to regulate chromatin remodeling and thus modulate

transcription. This could be mediated through the interaction between TFPT and INO80E. This is a particularly attractive hypothesis considering the well-established role of SUMOylation in regulating transcription factor activity through modulating assembly of multiprotein complexes.

A recent study found that knockdown of the INO80 ATPase, as well as the other INO80 complex subunits INO80E and TFPT, resulted in embryonic stem cell differentiation as measured by an ESC-specific reporter assay as well as loss of ESC morphology [149]. Additionally, knockdown of INO80E in mouse embryonic stem cells resulted in significant gene expression changes in >1200 genes as measured by a DNA microarray [149]. Interestingly, the INO80 complex was found to regulate the key pluripotency factor Sox2 and occupy multiple locations on the Sox2 gene in mouse embryonic stem cells. Additionally, a YY1 binding site identified by the ENCODE project [150] also overlaps with a TFPT DNA target sequence identified previously [8] in the orthologous region of the human genome. This data points to a potential role of both TFPT and YY1 working redundantly to recruit the INO80 complex to the Sox2 locus to regulate embryonic stem cell maintenance.

5.5. Methods

SUMO binding assay on human proteome microarrays

SUMO binding experiments were conducted with 4 probes representing the SUMO1, SUMO2 and the SUMO1 and SUMO2 trimers, all in triplicate. Protein chips were incubated overnight in blocking buffer buffer (20 mM HEPES-KOH pH 7.3, 100 mM potassium

acetate, 2 mM magnesium acetate, 1 mM EGTA, 0.05% Tween-20, 2% BSA) at 4° C. Chips were then incubated with SUMO1 or SUMO2 monomer or trimer in assay buffer (20 mM HEPES-KOH pH 7.3, 100 mM potassium acetate, 2 mM magnesium acetate, 1 mM EGTA, 0.05% Tween-20) for 1 hr at room temperature. Following incubation with SUMO, chips were washed 3X with assay buffer and 3X with PBS, followed by incubation with DyLight 549-anti-SUMO1 (21C7) or DyLight 649-anti-SUMO2 (8A2). Negative control experiments were done in parallel using antibodies only without SUMO. Chips were then washed 3X with 1X TBS with 0.05% Tween-20, 1X with water, then spun to dry and scanned.

Immunoprecipitation and Immunoblot

HEK 293T cells were transfected with Fugene 6 (Promega) and HCT 116 cells were transfected with Fugene HD (Promega). Both cell types were plated in 6-well plates, and harvested 24-48 hours following transfection. In order to detect protein SUMOylation, cells were washed with 1X PBS, then lysed in SUMO IP lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.5% sodium deoxycholate, 20 mM N-ethyl-maleimide, 1% SDS, and Roche protease inhibitor cocktail) for 10 minutes at 4° C. Then SUMO IP bind buffer was added to dilute lysate 1:10 and lysates were transferred to cold microcentrifuge tubes. Cells were then sonicated on ice and centrifuged at 15,000 g at 4° C for 10 minutes. Soluble cell lysates were incubated with dynabeads (Life Technologies) pre-bound to either anti-Flag (F1804, Sigma), anti-myc (R95025, Life Technologies), or anti-V5 (R9605, Life Technologies) antibodies for 2 hours at 4° C and washed 3 times with SUMO IP wash buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 2mM EDTA, 1% Triton X-100,

0.5% sodium deoxycholate, and 0.1% SDS). Bound proteins were eluted with 1X LDS-sample buffer (NP0008, Life Technologies) with 5% beta-mercaptoethanol. In order to look at protein interactions between INO80E and TFPT by co-immunoprecipitation, cells were transfected and harvested as above but lysed in INO80 lysis buffer (40 mM HEPES-NaOH pH 7.9, 0.3 M NaCl, 0.2% Triton X-100, and 10% glycerol), followed by centrifugation. Lysate supernatants were then bound to dynabeads linked to anti-FLAG or anti-V5 for 2 hours at 4° C and washed 3X with INO80 wash buffer (40 mM HEPES-NaOH 7.9, 0.25 M NaCl, 0.2% Triton X-100, and 10% glycerol). In both immunoprecipitation to detect SUMOylated proteins and coimmunoprecipitation experiments to detect interacting proteins, samples were resolved on 4-12% NuPage Bis-Tris gels (Life Technologies) in MES buffer and subject to immunoblotting using HRP-anti-V5 and HRP-anti-myc antibodies (Life Technologies).

Recombinant proteins

GST, GST-tagged SUMO1, SUMO2, and SUMO1 and SUMO2 trimers were expressed in bacteria and GST-INO80E was expressed in yeast. All GST-tagged proteins were purified by affinity chromatography on glutathione sepharose 4B beads (GE Healthcare) according to manufacturer's instructions.

In vitro binding assays

Recombinant GST, GST-tagged SUMO1, SUMO2, or SUMO2 trimer were diluted into 100 ul of 1X PBS, 0.05% Tween 20 and incubated with either glutathione-sepharose or alternatively glutathione-coated 96-well plates (Pierce Biotechnology). Following overnight incubation at 4 C, wells were blocked for 1 hr at room temperature with 2% bovine serum albumin in assay buffer (20 mM HEPES-KOH pH 7.3, 100 mM potassium acetate, 2 mM magnesium acetate, 1 mM EGTA, 0.05% Tween-20). INO80E and TFPT were produced by *in vitro* transcription and translation in rabbit reticulocyte lysate in the presence of [^{35}S] methionine according to manufacturer's instructions (Promega). *In vitro* translated proteins (10 µl) were diluted into 100 µl of assay buffer, when appropriate, incubated with SUMO1 or SUMO2 and the SUMO E1 and E2 enzymes, then incubated with the immobilized GST-tagged proteins for 1 hr at room temperature. Unbound proteins were removed by washing and bound proteins were eluted with SDS sample buffer and resolved by SDS-PAGE and autoradiography. Equal loading of immobilized proteins was verified by staining with SimplyBlue SafeStain (Life Technologies). Quantitative binding was determined by scanning densitometry measurements of radiolabeled protein relative to the amount of binding partner protein as indicated by Coomassie stain.

In vitro SUMOylation assays

TFPT was produced by *in vitro* transcription and translation in rabbit reticulocute lysate in the presence of [^{35}S] methionine according to manufacturer's instructions (Promega).

Radiolabeled protein was then incubated with our without SUMO1 or SUMO2 and the SUMO E1 and E2 enzymes for 1 hour at 37° C to yield a mixture of SUMOylated and unmodified TFPT for subsequent binding experiments.

Protein Microarray Data Analysis

To identify positive hits for each chip, first bad spots were removed, such as spots in a damaged region of the chip, as measured by a coefficient of variation (CV) value >1.5. A background correction was then applied, defined as the signal intensity (SI) value of each spot as the odds ratio of the foreground median divided by the background median. If a spot has a weak signal, its foreground median and background median will be close, thus its SI will be ~1. To normalize the signal, we suppose the true hits are rare and almost evenly dispersed in each block, thus we force each block on a chip to have a median SI of one. To be considered a positive hit, the duplicate spots of each protein must both have signal intensities (foreground/background ratio) 5 SD above the mean value. Additionally, positive hits must occur in 2 out of 3 triplicate experiments to be considered positive. Positive hits also identified in negative control chips were removed.

Network Motif Analysis

Protein microarray data from both SUMO-binding experiments and SUMOylation experiments [143] was integrated with protein-protein interaction data from publicly available databases, including BioGRID, DIP, MIPS, IntAct and HPRD and network motifs

were identified and analyzed as previously described [144]. The Z-value of a motif was calculated as the difference of its observed occurrence in a real network and its averaged occurrence in several hundred random networks, normalized with the standard deviation.

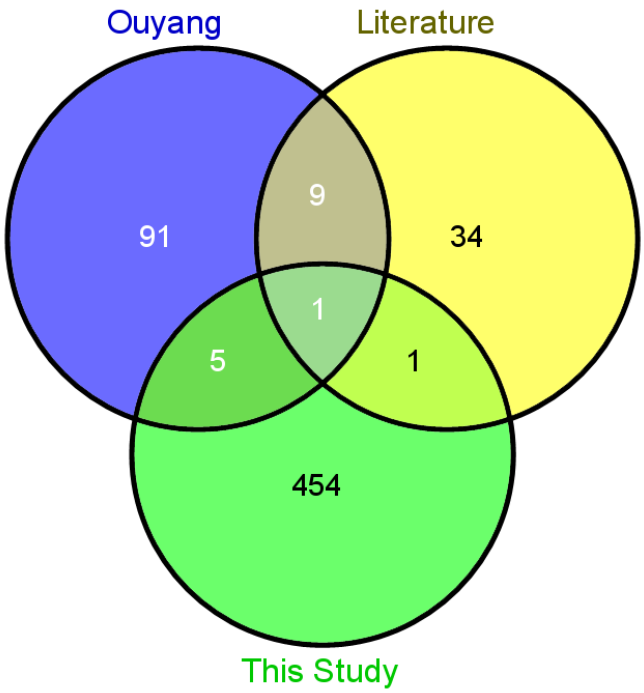

Figure 10. Venn diagram of SUMO-binding proteins identified here and in previous work. This illustrates the overlap between our dataset compared to proteins identified by Ouyang et al. [112] and other SUMO-binding proteins previously identified in the literature, which consists of a list of SUMO-binding proteins compiled by Zhao et al. [146] along with additional proteins that we compiled from more recent publications.

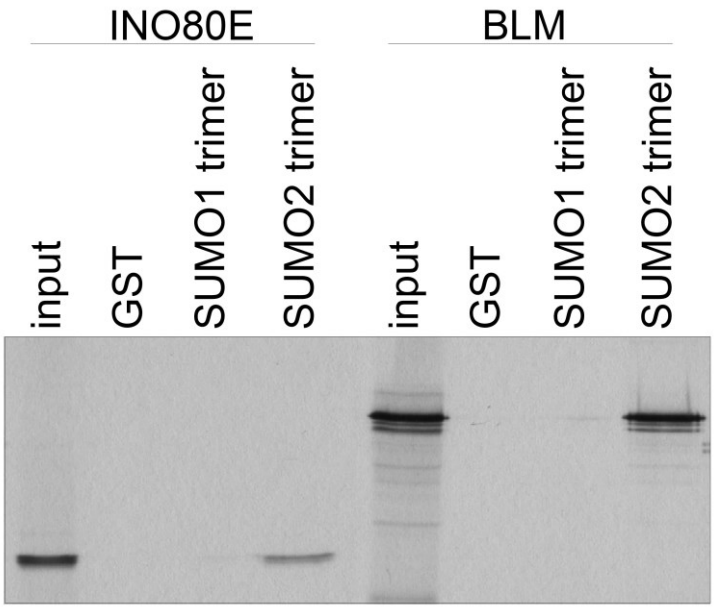

Figure 11. TFPT specifically binds the SUMO2 trimer *in vitro.* Results of *in vitro* binding assay showing specific binding of INO80E to the SUMO2 trimer. BLM, a known SUMO2 trimer binding protein [86] is shown as a positive control.

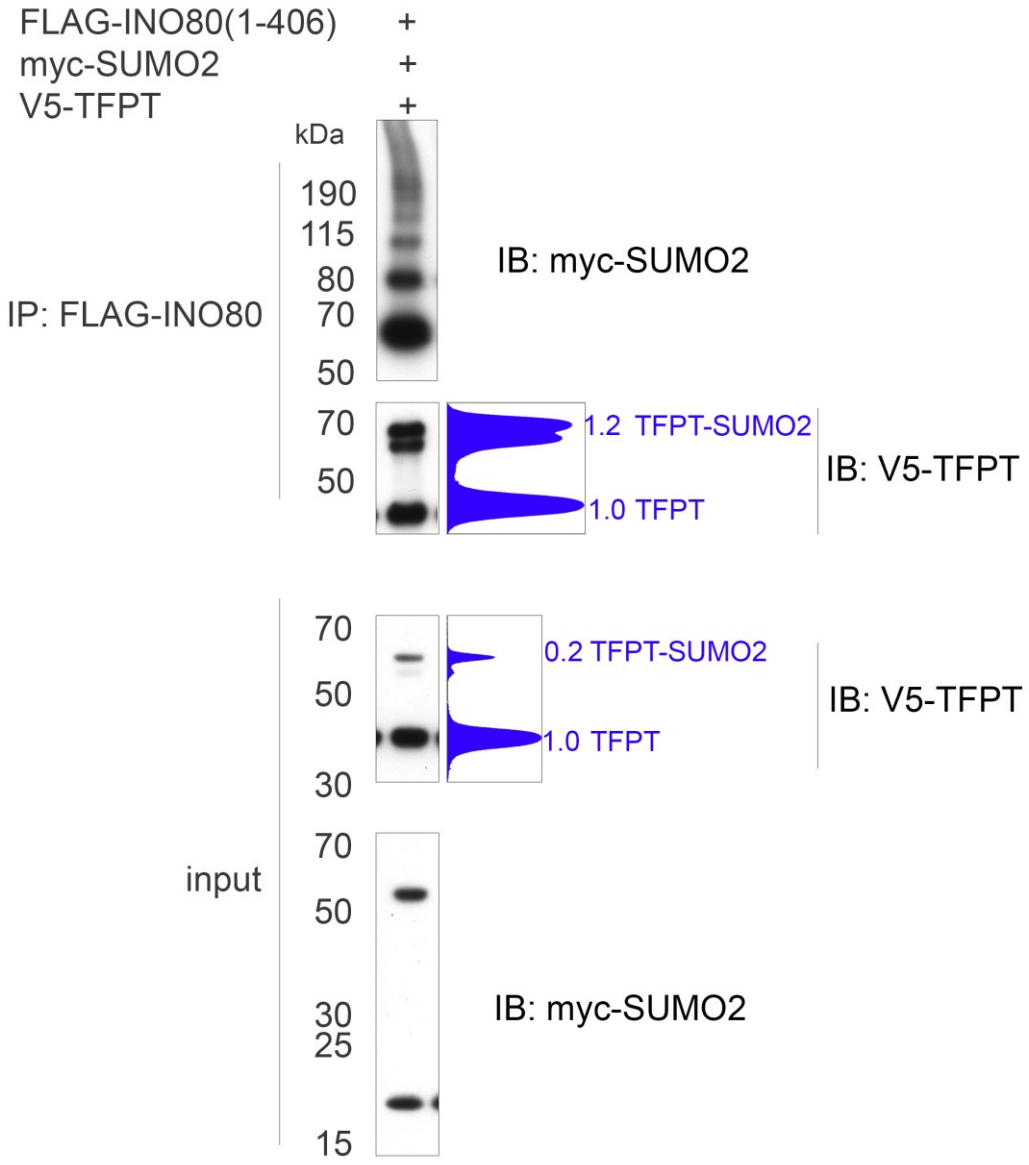

Figure 13. The INO80 ATPase N-terminal domain binds SUMOylated TFPT. Results of an *in vivo* binding assay. HCT116 cells were co-transfected with an N-terminal fragment of INO80 (1-406), SUMO2 and TFPT. We immunoprecipitated FLAG-tagged INO80 and immunoblotted with anti-V5 to detect both SUMO2-modified and unmodified forms of TFPT

References

[1] Zhu, H., Bilgin, M., Bangham, R., Hall, D.*, et al.*, Global analysis of protein activities using proteome chips. *Science* 2001, *293*, 2101-2105.

[2] Zhu, H., Hu, S., Jona, G., Zhu, X.*, et al.*, Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. *Proc Natl Acad Sci U S A* 2006, *103*, 4011-4016.

[3] Chen, C. S., Korobkova, E., Chen, H., Zhu, J.*, et al.*, A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. *Nat. Methods* 2008, *5*, 69-74.

[4] Popescu, S. C., Popescu, G. V., Bachan, S., Zhang, Z.*, et al.*, Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. *Proc Natl Acad Sci U S A* 2007, *104*, 4730-4735.

[5] Lueking, A., Possling, A., Huber, O., Beveridge, A.*, et al.*, A nonredundant human protein chip for antibody screening and serum profiling. *Mol. Cell. Proteomics* 2003, *2*, 1342-1349.

[6] Hu, S., Li, Y., Liu, G., Song, Q.*, et al.*, A protein chip approach for high-throughput antigen identification and characterization. *Proteomics* 2007, *7*, 2151-2161.

[7] Song, Q., Liu, G., Hu, S., Zhang, Y.*, et al.*, Novel autoimmune hepatitis-specific autoantigens identified using protein microarray technology. *J Proteome Res* 2010, *9*, 30-39.

[8] Hu, S., Xie, Z., Onishi, A., Yu, X.*, et al.*, Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* 2009, *139*, 610-622.

[9] Hu, S., Wan, J., Su, Y., Song, Q.*, et al.*, DNA methylation presents distinct binding sites for human transcription factors. *Elife* 2013, *2*, e00726.

[10] Jeong, J. S., Jiang, L., Albino, E., Marrero, J.*, et al.*, Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. *Mol. Cell. Proteomics* 2012, *11*, O111 016253.

[11] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 1999, *19*, 1720-1730.

[12] Kopf, E., Zharhary, D., Antibody arrays--an emerging tool in cancer proteomics. *The international journal of biochemistry & cell biology* 2007, *39*, 1305-1317.

[13] Zhu, H., Snyder, M., Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* 2001.

[14] Smith, M., Jona, G., Ptacek, J., Devgan, J.*, et al.*, Global analysis of protein function using protein function using protein microarrays. *Mech Age Dev* 2005.

[15] Xie, Z., Hu, S., Qian, J., Blackshaw, S., Zhu, H., Systematic characterization of protein-DNA interactions. *Cell. Mol. Life Sci.* 2011, *68*, 1657-1668.

[16] Roger P, E., Multi-analyte immunoassay. *Journal of Pharmaceutical and Biomedical Analysis* 1989, *7*, 155-168.

[17] Bussow, K., Cahill, D., Nietfeld, W., Bancroft, D.*, et al.*, A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* 1998, *26*, 5007-5008.

[18] Ge, H., UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res.* 2000, *28*, e3.

[19] MacBeath, G., Schreiber, S., Printing Proteins as Microarrays for High-Throughput Function Determination. *Nature* 2000.

[20] Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S.*, et al.*, Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proc Natl Acad Sci U S A* 2000, *97*, 10113-10119.

[21] Zhu, H., Klemic, J. F., Chang, S., Bertone, P.*, et al.*, Analysis of yeast protein kinases using protein chips. *Nat. Genet.* 2000, *26*, 283-289.

[22] Haab, B. B., Dunham, M. J., Brown, P. O., Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol* 2001, *2*, RESEARCH0004.

[23] Arenkov, P., Kukhtin, A., Gemmell, A., Voloshchuk, S.*, et al.*, Protein microchips: use for immunoassay and enzymatic reactions. *Anal. Biochem.* 2000, *278*, 123-131.

[24] Bieri, C., Ernst, O. P., Heyse, S., Hofmann, K. P., Vogel, H., Micropatterned immobilization of a G protein-coupled receptor and direct detection of G protein activation. *Nat. Biotechnol.* 1999, *17*, 1105-1108.

[25] Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Global Analysis of Protein Activities Using Proteome Chips. *Science* 2001.

[26] Feilner, T., Hultschig, C., Lee, J., Meyer, S.*, et al.*, High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. *Mol. Cell. Proteomics* 2005, *4*, 1558-1568.

[27] Chen, C.-S., Zhu, H., Protein microarrays. *BioTechniques* 2006, *40*, 423, 425, 427 passim.

[28] Smith, M. G., Jona, G., Ptacek, J., Devgan, G.*, et al.*, Global analysis of protein function using protein microarrays. *Mech Ageing Dev* 2005, *126*, 171-175.

[29] Poetz, O., Schwenk, J. M., Kramer, S., Stoll, D.*, et al.*, Protein microarrays: catching the proteome. *Mech Ageing Dev* 2005, *126*, 161-170.

[30] Hall, D. A., Zhu, H., Zhu, X., Royce, T.*, et al.*, Regulation of gene expression by a metabolic enzyme. *Science* 2004, *306*, 482-484.

[31] Zhu, J., Gopinath, K., Murali, A., Yi, G.*, et al.*, RNA-binding proteins that inhibit RNA virus infection. *Proc Natl Acad Sci U S A* 2007, *104*, 3129-3134.

[32] Ptacek, J., Devgan, G., Michaud, G., Zhu, H.*, et al.*, Global analysis of protein phosphorylation in yeast. *Nature* 2005, *438*, 679-684.

[33] Lin, Y. Y., Lu, J. Y., Zhang, J., Walter, W.*, et al.*, Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell* 2009, *136*, 1073-1084.

[34] Lu, J.-Y., Lin, Y.-Y., Qian, J., Tao, S.-C.*, et al.*, Functional Dissection of a HECT Ubiquitin E3 Ligase. *Molecular & Cellular Proteomics* 2007, *7*, 35-45.

[35] Lu, J. Y., Lin, Y. Y., Sheu, J. C., Wu, J. T.*, et al.*, Acetylation of yeast AMPK controls intrinsic aging independently of caloric restriction. *Cell* 2011, *146*, 969-979.

[36] Oh, Y.-H., Hong, M.-Y., Jin, Z., Lee, T.*, et al.*, Chip-based analysis of SUMO (small ubiquitin-like modifier) conjugation to a target protein. *Biosensors & bioelectronics* 2007, *22*, 1260-1267.

[37] Jeong, J. S., Rho, H.-s., Zhu, H., A functional protein microarray approach to characterizing posttranslational modifications on lysine residues. *Methods in molecular biology (Clifton, NJ)* 2011, *723*, 213-223.

[38] Knezevic, V., Leethanakul, C., Bichsel, V. E., Worth, J. M.*, et al.*, Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* 2001, *1*.

[39] Schnack, C., Hengerer, B., Gillardon, F., Identification of novel substrates for Cdk5 and new targets for Cdk5 inhibitors using high-density protein microarrays. *Proteomics* 2008, *8*, 1980-1986.

[40] Tarrant, M. K., Rho, H. S., Xie, Z., Jiang, Y. L.*, et al.*, Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. *Nat. Chem. Biol.* 2012, *8*, 262-269.

[41] Nikitin, P. A., Luftig, M. A., The DNA damage response in viral-induced cellular transformation. *Br. J. Cancer* 2012, *106*, 429-435.

[42] Zhu, J., Liao, G., Shan, L., Zhang, J.*, et al.*, Protein array identification of substrates of the Epstein-Barr virus protein kinase BGLF4. *J. Virol.* 2009, *83*, 5219-5231.

[43] Gupta, R., Kus, B., Fladd, C., Wasmuth, J.*, et al.*, Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast. *Mol. Syst. Biol.* 2007, *3*, 116.

[44] Persaud, A., Alberts, P., Amsen, E. M., Xiong, X.*, et al.*, Comparison of substrate specificity of the ubiquitin ligases Nedd4 and Nedd4-2 using proteome arrays. *Mol. Syst. Biol.* 2009, *5*, 333.

[45] Thao, S., Chen, C. S., Zhu, H., Escalante-Semerena, J. C., Nepsilon-lysine acetylation of a bacterial transcription factor inhibits Its DNA-binding activity. *PloS one* 2010, *5*, e15123.

[46] Foster, M. W., Forrester, M. T., Stamler, J. S., A protein microarray-based analysis of S-nitrosylation. *Proc Natl Acad Sci U S A* 2009, *106*, 18948-18953.

[47] Zheng, D., Wan, J., Cho, Y. G., Wang, L.*, et al.*, Comparison of humoral immune responses to Epstein-Barr virus and Kaposi's sarcoma-associated herpesvirus using a viral proteome microarray. *J. Infect. Dis.* 2011, *204*, 1683-1691.

[48] Li, B., Jiang, L., Song, Q., Yang, J.*, et al.*, Protein microarray for profiling antibody responses to Yersinia pestis live vaccine. *Infect. Immun.* 2005, *73*, 3734-3739.

[49] Vigil, A., Chen, C., Jain, A., Nakajima-Sasaki, R.*, et al.*, Profiling the humoral immune response of acute and chronic Q fever by protein microarray. *Mol. Cell. Proteomics* 2011, *10*, M110 006304.

[50] Doolan, D. L., Mu, Y., Unal, B., Sundaresh, S.*, et al.*, Profiling humoral immune responses to P. falciparum infection with protein microarrays. *Proteomics* 2008, *8*, 4680-4694.

[51] Chen, C.-S., Sullivan, S., Anderson, T., Tan, A. C.*, et al.*, Identification of novel serological biomarkers for inflammatory bowel disease using Escherichia coli proteome chip. *Molecular & Cellular Proteomics* 2009, *8*, 1765-1776.

[52] Liang, L., Tan, X., Juarez, S., Villaverde, H.*, et al.*, Systems biology approach predicts antibody signature associated with Brucella melitensis infection in humans. *J Proteome Res* 2011, *10*, 4813-4824.

[53] Hu, C. J., Song, G., Huang, W., Liu, G. Z.*, et al.*, Identification of new autoantigens for primary biliary cirrhosis using human proteome microarrays. *Mol. Cell. Proteomics* 2012, *11*, 669-680.

[54] Hu, S., Vissink, A., Arellano, M., Roozendaal, C.*, et al.*, Identification of autoantibody biomarkers for primary Sjogren's syndrome using protein microarrays. *Proteomics* 2011, *11*, 1499-1507.

[55] Anderson, K. S., Sibani, S., Wallstrom, G., Qiu, J.*, et al.*, Protein microarray signature of autoantibody biomarkers for the early detection of breast cancer. *Journal of proteome research* 2011, *10*, 85-96.

[56] Tao, S. C., Li, Y., Zhou, J., Qian, J.*, et al.*, Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* 2008, *18*, 761-769.

[57] Orenes-Piñero, E., Barderas, R., Rico, D., Casal, J. I.*, et al.*, Serum and tissue profiling in bladder cancer combining protein and tissue arrays. *Journal of proteome research* 2010, *9*, 164-173.

[58] Petricoin, E. F., 3rd, Espina, V., Araujo, R. P., Midura, B.*, et al.*, Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. *Cancer Res.* 2007, *67*, 3431-3440.

[59] Luevano, M., Bernard, H. U., Barrera-Saldana, H. A., Trevino, V.*, et al.*, High-throughput profiling of the humoral immune responses against thirteen human papillomavirus types by proteome microarrays. *Virology* 2010, *405*, 31-40.

[60] Arvin, A., Abendroth, A., in: Arvin, A., Campadelli-Fiume, G., Mocarski, E., Moore, P. S.*, et al.* (Eds.), *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, Cambridge 2007.

[61] Li, R., Wang, L., Liao, G., Guzzo, C. M.*, et al.*, SUMO binding by the Epstein-Barr virus protein kinase BGLF4 is crucial for BGLF4 function. *J. Virol.* 2012, *86*, 5412-5421.

[62] Margarit, I., Bonacci, S., Pietrocola, G., Rindi, S.*, et al.*, Capturing host-pathogen interactions by protein microarrays: identification of novel streptococcal proteins binding to human fibronectin, fibrinogen, and C4BP. *FASEB J.* 2009, *23*, 3100-3112.

[63] Komarova, A. V., Combredet, C., Meyniel-Schicklin, L., Chapelle, M.*, et al.*, Proteomic analysis of virus-host interactions in an infectious context using recombinant viruses. *Mol. Cell. Proteomics* 2011, *10*, M110 007443.

[64] Krishnan, M. N., Ng, A., Sukumaran, B., Gilfoy, F. D.*, et al.*, RNA interference screen for human genes associated with West Nile virus infection. *Nature* 2008, *455*, 242-245.

[65] Brass, A. L., Huang, I. C., Benita, Y., John, S. P.*, et al.*, The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* 2009, *139*, 1243-1254.

[66] Uetz, P., Dong, Y. A., Zeretzke, C., Atzler, C.*, et al.*, Herpesviral protein networks and their interaction with the human proteome. *Science* 2006, *311*, 239-242.

[67] Pfefferle, S., Schopf, J., Kogl, M., Friedel, C. C.*, et al.*, The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog* 2011, *7*, e1002331.

[68] Shamay, M., Liu, J., Li, R., Liao, G.*, et al.*, A protein array screen for Kaposi's sarcoma-associated herpesvirus LANA interactors links LANA to TIP60, PP2A activity, and telomere shortening. *J. Virol.* 2012, *86*, 5179-5191.

[69] Parsons, D. W., Li, M., Zhang, X., Jones, S.*, et al.*, The genetic landscape of the childhood cancer medulloblastoma. *Science* 2011, *331*, 435-439.

[70] Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J.*, et al.*, Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012, *366*, 883-892.

[71] Murphy, M. A., O'Leary, J. J., Cahill, D. J., Assessment of the humoral immune response to cancer. *Journal of proteomics* 2012, *75*, 4573-4579.

[72] Fukata, M., Fukata, Y., Adesnik, H., Nicoll, R. A., Bredt, D. S., Identification of PSD-95 palmitoylating enzymes. *Neuron* 2004, *44*, 987-996.

[73] Geiss-Friedlander, R., Melchior, F., Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* 2007, *8*, 947-956.

[74] Martin, S., Wilkinson, K. A., Nishimune, A., Henley, J. M., Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction. *Nat. Rev. Neurosci.* 2007, *8*, 948-959.

[75] Meluh, P. B., Koshland, D., Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Molecular biology of the cell* 1995, *6*, 793-807.

[76] Matunis, M. J., Coutavas, E., Blobel, G., A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* 1996, *135*, 1457-1470.

[77] Okura, T., Gong, L., Kamitani, T., Wada, T.*, et al.*, Protection against Fas/APO-1- and tumor necrosis factor-mediated cell death by a novel protein, sentrin. *J. Immunol.* 1996, *157*, 4277-4281.

[78] Boddy, M. N., Howe, K., Etkin, L. D., Solomon, E., Freemont, P. S., PIC 1, a novel ubiquitin-like protein which interacts with the PML component of a multiprotein complex that is disrupted in acute promyelocytic leukaemia. *Oncogene* 1996, *13*, 971-982.

[79] Shen, Z., Pardington-Purtymun, P. E., Comeaux, J. C., Moyzis, R. K., Chen, D. J., UBL1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins. *Genomics* 1996, *36*, 271-279.

[80] Mahajan, R., Delphin, C., Guan, T., Gerace, L., Melchior, F., A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* 1997, *88*, 97-107.

[81] Ayaydin, F., Dasso, M., Distinct in vivo dynamics of vertebrate SUMO paralogues. *Molecular biology of the cell* 2004, *15*, 5208-5218.

[82] Matic, I., van Hagen, M., Schimmel, J., Macek…, B., In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Molecular & cellular …* 2008.

[83] Saitoh, H., Hinchey, J., Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* 2000, *275*, 6252-6258.

[84] Meulmeester, E., Kunze, M., Hsiao, H. H., Urlaub, H., Melchior, F., Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. *Mol. Cell* 2008, *30*, 610-619.

[85] Azuma, Y., Arnaoutov, A., Dasso, M., SUMO-2/3 regulates topoisomerase II in mitosis. *J. Cell Biol.* 2003, *163*, 477-487.

[86] Zhu, J., Zhu, S., Guzzo, C. M., Ellis, N. A.*, et al.*, Small ubiquitin-related modifier (SUMO) binding determines substrate recognition and paralog-selective SUMO modification. *J. Biol. Chem.* 2008, *283*, 29405-29415.

[87] Johnson, E., Gupta, A., An E3-like factor that promotes SUMO conjugation to the yeast septins. *Cell* 2001.

[88] Sachdev, S., Bruhn, L., Sieber, H., Pichler, A.*, et al.*, PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev.* 2001, *15*, 3088-3103.

[89] Kotaja, N., Karvonen, U., Janne, O. A., Palvimo, J. J., PIAS Proteins Modulate Transcription Factors by Functioning as SUMO-1 Ligases. *Mol. Cell. Biol.* 2002, *22*, 5222-5234.

[90] Kahyo, T., Nishida, T., Yasuda, H., Involvement of PIAS1 in the Sumoylation of Tumor Suppressor p53. *Mol. Cell* 2001.

[91] Takahashi, Y., Toh-e, A., Kikuchi, Y., A novel factor required for the SUMO1/Smt3 conjugation of yeast septins. *Gene* 2001, *275*, 223-231.

[92] Deng, Z., Wan, M., Sui, G., PIASy-mediated sumoylation of Yin Yang 1 depends on their interaction but not the RING finger. *Mol. Cell. Biol.* 2007, *27*, 3780-3792.

[93] Weger, S., Hammer, E., Heilbronn, R., Topors acts as a SUMO-1 E3 ligase for p53 in vitro and in vivo. *FEBS Lett.* 2005, *579*, 5007-5012.

[94] Braschi, E., Zunino, R., McBride, H. M., MAPL is a new mitochondrial SUMO E3 ligase that regulates mitochondrial fission. *EMBO Rep* 2009, *10*, 748-754.

[95] Li, W., Bengtson, M., Ulbrich, A., Matsuda, A., … functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 …. *PloS one* 2008.

[96] Takahashi, H., Hatakeyama, S., Saitoh, H., Nakayama, K. I., Noncovalent SUMO-1 binding activity of thymine DNA glycosylase (TDG) is required for its SUMO-1 modification and colocalization with the promyelocytic leukemia protein. *J. Biol. Chem.* 2005, *280*, 5611-5621.

[97] Ouyang, J., Gill, G., SUMO engages multiple corepressors to regulate chromatin structure and transcription. *Epigenetics : official journal of the DNA Methylation Society* 2009, *4*, 440-444.

[98] Girdwood, D., Bumpass, D., Vaughan, O. A., Thain, A., *et al.*, P300 transcriptional repression is mediated by SUMO modification. *Mol. Cell* 2003, *11*, 1043-1054.

[99] Yang, S. H., Sharrocks, A. D., SUMO promotes HDAC-mediated transcriptional repression. *Mol. Cell* 2004, *13*, 611-617.

[100] Matunis, M., Wu, J., Blobel, G., SUMO-1 Modification and Its Role in Targeting the Ran GTPase-activating Protein, RanGAP1, to the Nuclear Pore Complex. *The Journal of cell biology* 1998, *140*, 499.

[101] Pfander, B., Moldovan, G.-L., Sacher, M., Hoege, C., Jentsch, S., SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* 2005, *436*, 428-433.

[102] Minty, A., Dumont, X., Kaghad, M., Caput, D., Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif. *J. Biol. Chem.* 2000, *275*, 36316-36323.

[103] Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G., Chen, Y., Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci U S A* 2004, *101*, 14373-14378.

[104] Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P., Dikic, I., Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* 2006, *281*, 16117-16127.

[105] Danielsen, J. R., Povlsen, L. K., Villumsen, B. H., Streicher, W.*, et al.*, DNA damage-inducible SUMOylation of HERC2 promotes RNF8 binding via a novel SUMO-binding Zinc finger. *J. Cell Biol.* 2012, *197*, 179-187.

[106] Guzzo, C. M., Ringel, A., Cox, E., Uzoma, I.*, et al.*, Characterization of the SUMO-Binding Activity of the Myeloproliferative and Mental Retardation (MYM)-Type Zinc Fingers in ZNF261 and ZNF198. *PloS one* 2014, *9*, e105271.

[107] O'Rourke, J. G., Gareau, J. R., Ochaba, J., Song, W.*, et al.*, SUMO-2 and PIAS1 modulate insoluble mutant huntingtin protein accumulation. *Cell reports* 2013, *4*, 362-375.

[108] Krumova, P., Meulmeester, E., Garrido, M., Tirard, M.*, et al.*, Sumoylation inhibits alpha-synuclein aggregation and toxicity. *J. Cell Biol.* 2011, *194*, 49-60.

[109] Kho, C., Lee, A., Jeong, D., Oh, J. G.*, et al.*, SUMO1-dependent modulation of SERCA2a in heart failure. *Nature* 2011, *477*, 601-605.

[110] Wang, Q., Xia, N., Li, T., Xu, Y.*, et al.*, SUMO-specific protease 1 promotes prostate cancer progression and metastasis. *Oncogene* 2013, *32*, 2493-2498.

[111] Jackson, S. P., Durocher, D., Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* 2013, *49*, 795-807.

[112] Ouyang, J., Shi, Y., Valin, A., Xuan, Y., Gill, G., Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol. Cell* 2009, *34*, 145-154.

[113] Morrison, A. J., Shen, X., Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat. Rev. Mol. Cell Biol.* 2009, *10*, 373-384.

[114] Papamichos-Chronakis, M., Watanabe, S., Rando, O. J., Peterson, C. L., Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* 2011, *144*, 200-213.

[115] Jin, J., Cai, Y., Yao, T., Gottschalk, A. J.*, et al.*, A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *J. Biol. Chem.* 2005, *280*, 41207-41212.

[116] Chen, L., Cai, Y., Jin, J., Florens, L.*, et al.*, Subunit organization of the human INO80 chromatin remodeling complex: an evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling. *J. Biol. Chem.* 2011, *286*, 11283-11289.

[117] Brambillasca, F., Mosna, G., Colombo, M., Rivolta…, A., Identification of a novel molecular partner of the E2A gene in childhood leukemia. *Leukemia* 1999.

[118] Irie, Y., Yamagata, K., Gan, Y., Miyamoto, K.*, et al.*, Molecular cloning and characterization of Amida, a novel protein which interacts with a neuron-specific immediate early gene product arc, contains novel nuclear localization signals, and causes cell death in cultured cells. *The Journal of biological chemistry* 2000, *275*, 2647-2653.

[119] Franchini, C., Fontana, F., Minuzzo, M., Babbio, F., Privitera, E., Apoptosis promoted by up-regulation of TFPT (TCF3 fusion partner) appears p53 independent, cell type restricted and cell density influenced. *Apoptosis* 2006, *11*, 2217-2224.

[120] Jiang, F., Ai, J., Xiao, W., Wang, Z., FB1, an E2A fusion partner in childhood leukemia, interacts with U19/EAF2 and inhibits its transcriptional activity. *Cancer Lett.* 2007, *253*, 265-272.

[121] Jin, J., Cai, Y., Yao, T., Gottschalk, A. J.*, et al.*, A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *The Journal of biological chemistry* 2005, *280*, 41207-41212.

[122] Hu, S., Xie, Z., Onishi, A., Yu, X.*, et al.*, Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* 2009, *139*, 610-622.

[123] Bloomer, W. A. C., VanDongen, H. M. A., VanDongen, A. M. J., Activity-regulated cytoskeleton-associated protein Arc/Arg3.1 binds to spectrin and associates with nuclear promyelocytic leukemia (PML) bodies. *Brain Res.* 2007, *1153*, 20-33.

[124] Irie, Y., Molecular Cloning and Characterization of Amida, a Novel Protein Which Interacts with a Neuron-specific Immediate Early Gene Product Arc, Contains Novel Nuclear Localization Signals, and Causes Cell Death in Cultured Cells. *J. Biol. Chem.* 2000, *275*, 2647-2653.

[125] Golebiowski, F., Matic, I., Tatham, M., Cole, C., Hay, R., System-Wide Changes to SUMO Modifications in Response to Heat Shock. *Science's STKE* 2009.

[126] Tammsalu, T., Matic, I., Jaffray, E. G., Ibrahim, A. F. M.*, et al.*, Proteome-wide identification of SUMO2 modification sites. *Science signaling* 2014, *7*, rs2.

[127] Rudra, D., deRoos, P., Chaudhry, A., Niec, R. E.*, et al.*, Transcription factor Foxp3 and its protein partners form a complex regulatory network. *Nat. Immunol.* 2012, *13*, 1010-1019.

[128] Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S.*, et al.*, A draft map of the human proteome. *Nature* 2014, *509*, 575-581.

[129] Ivanov, A. V., Peng, H., Yurchenko, V., Yap, K. L.*, et al.*, PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell* 2007, *28*, 823-837.

[130] Stielow, B., Sapetschnig, A., Krüger, I., Kunert, N.*, et al.*, Identification of SUMO-dependent chromatin-associated transcriptional repression components by a genome-wide RNAi screen. *Mol. Cell* 2008, *29*, 742-754.

[131] Lee, Y. I., Giovinazzo, D., Kang, H. C., Lee, Y.*, et al.*, Protein microarray characterization of the S-nitrosoproteome. *Mol. Cell. Proteomics* 2014, *13*, 63-72.

[132] Barry, G., Briggs, J. A., Vanichkina, D. P., Poth, E. M.*, et al.*, The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry* 2014, *19*, 486-494.

[133] Donnelly, C. J., Zhang, P. W., Pham, J. T., Heusler, A. R.*, et al.*, RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* 2013, *80*, 415-428.

[134] Oh, Y., Chung, K. C., UHRF2, a ubiquitin E3 ligase, acts as a small ubiquitin-like modifier E3 ligase for zinc finger protein 131. *J. Biol. Chem.* 2013, *288*, 9102-9111.

[135] Liang, Q., Deng, H., Li, X., Wu, X.*, et al.*, Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J. Immunol.* 2011, *187*, 4754-4763.

[136] Garcia-Gutierrez, P., Juarez-Vicente, F., Gallardo-Chamizo, F., Charnay, P., Garcia-Dominguez, M., The transcription factor Krox20 is an E3 ligase that sumoylates its Nab coregulators. *EMBO Rep* 2011, *12*, 1018-1023.

[137] dos Santos, M. T., Trindade, D. M., Goncalves Kde, A., Bressan, G. C.*, et al.*, Human stanniocalcin-1 interacts with nuclear and cytoplasmic proteins and acts as a SUMO E3 ligase. *Mol Biosyst* 2011, *7*, 180-193.

[138] Chu, Y., Yang, X., SUMO E3 ligase activity of TRIM proteins. *Oncogene* 2011, *30*, 1108-1116.

[139] Gill, G., SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev.* 2004, *18*, 2046-2059.

[140] Reverter, D., Lima, C. D., Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* 2005, *435*, 687-692.

[141] Capili, A. D., Lima, C. D., Structure and analysis of a complex between SUMO and Ubc9 illustrates features of a conserved E2-Ubl interaction. *J. Mol. Biol.* 2007, *369*, 608-618.

[142] Keusekotten, K., Bade, V. N., Meyer-Teschendorf, K., Sriramachandran, A. M.*, et al.*, Multivalent interactions of the SUMO-interaction motifs in RING finger protein 4 determine the specificity for chains of the SUMO. *Biochem. J.* 2014, *457*, 207-214.

[143] Uzoma, I. H., J.; Cox, E.; Guzzo, C.M.; Matunis, M.J.; Blackshaw, S.; Qian, J.; Zhu, H., SUMOylation Regulates MAPK Signaling. 2014.

[144] Hwang, W., Hackler, L., Jr., Wu, G., Ji, H.*, et al.*, Dynamics of regulatory networks in the developing mouse retina. *PloS one* 2012, *7*, e46521.

[145] Golebiowski, F., Matic, I., Tatham, M. H., Cole, C.*, et al.*, System-wide changes to SUMO modifications in response to heat shock. *Sci Signal* 2009, *2*, ra24.

[146] Zhao, Q., Xie, Y., Zheng, Y., Jiang, S.*, et al.*, GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res.* 2014, *42*, W325-330.

[147] Lyst, M. J., Nan, X., Stancheva, I., Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *EMBO J.* 2006, *25*, 5317-5328.

[148] Cai, Y., Jin, J., Yao, T., Gottschalk, A. J.*, et al.*, YY1 functions with INO80 to activate transcription. *Nat. Struct. Mol. Biol.* 2007, *14*, 872-874.

[149] Wang, L., Du, Y., Ward, J. M., Shimbo, T.*, et al.*, INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell stem cell* 2014, *14*, 575-591.

[150] Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R.*, et al.*, ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* 2013, *41*, D56-63.

CURRICULUM VITAE

Eric Cox Updated August 22, 2014

Date of Birth: May 9th, 1982. Santa Monica, CA

Education and Training

2014 PhD expected in Biochemistry, Cellular and Molecular Biology (BCMB), Johns Hopkins University School of Medicine, Baltimore, MD

2004 BS, General Biology, University of California, San Diego, La Jolla, CA

Professional Experience

2005-2007 Laboratory Technician, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

2003-2005 Laboratory Assistant, Department of Psychiatry, University of California, San Diego, La Jolla, CA

Fellowships and Awards

2010-2012 American Heart Association Predoctoral Fellowship

Teaching Experience

2010 Graduate Course Tutor, Pathways and Regulation, Biochemistry, Cellular and Molecular Biology Program Core Curriculum Couse, Johns Hopkins University School of Medicine, Baltimore, MD

2003 Teacher's Assistant, Biometry, Division of Biology, University of California San Diego, La Jolla, CA

Publications

1. Cox E, Uzoma I, Guzzo C, Jeong JS, Matunis MJ, Blackshaw S, Zhu H. Identification of SUMO E3 ligase specific substrates using the HuProt human proteome microarray. Methods in Molecular Biology (submitted)
2. Guzzo CM, Ringel A, Cox E, Uzoma I, Zhu H, Blackshaw S, Wolberger C, Matunis MJ. Characterization of the SUMO-binding activity of the myeloproliferative and mental retardation (MYM)-type zinc fingers in ZNF261 and ZNF198. PLoS One 9: e105271, 2014.

3. Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C, Xia S, Liu S, Lyu H, Ming GL, Wade H, Song H, Qian J, Zhu H. DNA methylation presents distinct binding sites for human transcription factors. Elife 2: e00726, 2013
4. Zhu H, Cox E, Qian J. Functional protein microarray as molecular decathlete: a versatile player in clinical proteomics. Proteomics Clin Appl. 6: 548-562, 2012
5. Lanier J, Quina LA, Eng SR, Cox E, Turner EE. Brn3a target recognition in embryonic sensory neurons. Dev Biol 302: 703-716, 2007
6. Cox E, Lanier J, Quina LA, Eng SR, Turner EE. Regulation of FGF10 by POU transcription factor Brn3a in the developing trigeminal ganglion. J Neurobiol. 66: 1075-1083, 2006

Abstracts and Presentations

1. **Cox E,** Guzzo, C, Hu J, Uzoma I, Jeong JS, Qian J, Matunis MJ, Blackshaw S, Zhu H. Identifying SUMO Binding Proteins by Human Proteome Microarray. Poster Presentation, National Technology Centers for Networks and Pathways — All Hands Meeting. Bethesda, MD. July 2012
2. **Cox E,** Guzzo, C, Hu J, Uzoma I, Jeong JS, Qian J, Matunis MJ, Blackshaw S, Zhu H. Identifying SUMO Binding Proteins by Human Proteome Microarray. Poster Presentation, SUMO, Ubiquitin, UBL proteins Fifth International Conference, Houston, TX. February 2012

Professional Memberships

American Society for Biochemistry and Molecular Biology